Do Sell-side Analyst Reports Have Investment Value?

Linying Lv*

First draft: January 2025. This draft: July 2025.

Abstract

This paper documents novel investment value in analyst report text. Using 1.2 million reports from 2000–2023, I embed narratives with large language models (LLMs) and fit machine learning (ML) forecasts of 12-month returns. Portfolios formed on the report narrative forecasts earn sizable and statistically significant performance that is incremental to analysts' numerical outputs and to a broad set of established factors and characteristic-based predictors. The effect is strongest after adverse contemporaneous news and is amplified for reports authored by skilled, experienced analysts. To open the black box, I apply a Shapley decomposition that attributes portfolio performance to distinct topics. Analysts' strategic outlook contributes the most to portfolio performance, especially forward-looking fundamental assessments. Beyond providing direct evidence that analyst narratives contain value-relevant assessments that diffuse into price over time, this study illustrates how interpretable LLM-plus-ML pipelines can scale and augment human judgment in investment decisions.

JEL Classification: G12, G14, G24

Keywords: Analyst Report, Investment Value, Text Alpha, Large Language Models, Explainable AI

*Olin Business School, Washington University in St. Louis, llyu@wustl.edu.

1 Introduction

Sell-side analysts convey perspectives about stocks through their research outputs. A large literature studies their numerical outputs—recommendations, earnings forecasts, and target prices—and documents that quantitative forecasts generate profitable trading signals (e.g., Womack, 1996; Brav and Lehavy, 2003; Jegadeesh et al., 2004; Birru et al., 2022; Farago et al., 2023). In contrast, much less is known about whether the accompanying narrative contains incremental investment value and, if so, where the value originates.

In this paper, I analyze 1.2 million analyst reports from Thomson Reuters Investext between 2000 and 2023 to test whether report narratives contain incremental investment value and to identify what drives that value within an integrated framework spanning information representation, relevance, and interpretation. LLMs are crucial here because report text is unstructured and high-dimensional: unlike traditional methods that compress documents into a single sentiment score, LLMs preserve multi-dimensional, context-aware structural representations. This richer representation allows me to study instances where narratives diverge from numerical forecasts and to distinguish genuine analyst insight from measurement error. It also mitigates the context-dependence and misclassification issues that hamper dictionary-based sentiment approaches, as highlighted by Loughran and McDonald (2011).

Methodologically, I proceed in three steps. First, I map analyst reports into a high-dimensional vector space using LLM transformer embeddings. These embeddings capture the LLM's deep linguistic understanding and semantic interpretation of the text, representing a sophisticated "thinking process" that converts textual input into structured output.[1] Second, I use machine learning models to extract value-relevant information to predict future 12-month stock returns and construct trading strategies based on these predictions. Finally, I conduct attribution analysis using Shapley value decomposition to identify which report components explain portfolio performances.

In initial predictive regressions, embeddings generated from the narrative text exhibit significant out-of-sample power to forecast 12-month stock returns. This predictability remains statistically and economically significant after controlling for analysts' quantitative outputs, including recommendation, earnings forecast, and target price revisions, and when using DGTW characteristic-adjusted returns.

To assess economic significance, I construct a monthly rebalanced portfolio sorted on these text-based return predictions. A self-financing trading strategy that buys stocks with the most positive predictions and shorts those with the most negative predictions generates

[1] This approach of text representation has gained traction in recent studies (Jha et al., 2020; Chen et al., 2022; Chen et al., 2023b; Lv, 2024; Siano, 2025).

an average monthly return of 1.04% and an of 68 basis points (t = 2.64) against the Fama and French (2018) model. To rigorously evaluate the incremental value, I benchmark the strategy against 94 fundamental-based factors from Gu et al. (2020) and 18 analyst-based factors from Chen and Zimmermann (2022). The narrative-based strategy's alpha is not subsumed by these established predictors, delivering high information ratios ranging from 0.73 to 1.41. This demonstrates that analyst narratives contain a distinct and substantial source of investment value.

Predictive power is not uniform across the cross-section. It is concentrated in large-cap, profitable, low book-to-market firms, and is stronger when reports are authored by analysts with higher historical forecast accuracy and greater experience. Consistent with this firm profile, factor regressions show the text-based strategy has significant negative exposures to the value and investment factors.

The timing of the effect further sharpens the interpretation. Conditioning on contemporaneous news, I find that predictability strengthens precisely after adverse signals: low earnings surprises, downgrades, and below-median earnings- or target-price revisions. This state dependence is hard to reconcile with a drift-only mechanism like standard post-announcement drift. It also cannot be reproduced by coarse sentiment sorting. Decile strategies formed on announcement returns, headline tone, body tone, or recommendation revisions do not deliver robust alphas, whereas the text-based forecast remains significant when sentiment variables enter directly as controls. Together, the cross-sectional and state-dependent results suggest a common mechanism: credible analysts' narratives contain valuable long-horizon prospects for profitable growth firms, especially when recent news is unfavorable, and this information diffuses into prices over time.

A natural question is which parts of analyst reports drive investment value. I address this using a Shapley value decomposition that attributes each report's predictive power to its constituent topics. The idea is simple: after partitioning each report into topics, I evaluate how portfolio performance changes across all possible topic subsets. Averaging marginal contributions over these subsets yields a fair estimate of each topic's share of the final outcome.

I first build a topic taxonomy that is both exclusive and economically meaningful. Using a modern large language model to draft a hierarchy and manual consolidation, I organize content into five meta-categories: (i) Company & Industry Overview, (ii) Financial Analysis, (iii) Strategic Outlook, (iv) Risk & Governance, and (v) Additional Content. I then fine-tune a BERT classifier to assign 52 million sentences to these categories. With sentences labeled, I construct topic-specific embeddings, generate return forecasts by topic, and compute Shapley values for each category's contribution to portfolio Sharpe ratio and returns.

The attribution analysis reveals a striking finding: the Strategic Outlook category, despite representing only 15% of report content, accounts for 41% of the portfolio's Sharpe ratio. In contrast, Financial Analysis-the most prevalent category-contributes only 16%, suggesting that financial data are more rapidly incorporated into prices. The remaining categories offer moderate to negligible contributions. Univariate portfolio sorts corroborate this attribution. A strategy trading solely on signals from the Strategic Outlook content generates a monthly return of 1.41% (Sharpe ratio $= 0.93$), substantially outperforming strategies based on any other single category (returns of $0.52\% - 0.58\%$; Sharpe ratios of $0.37 - 0.41$).

What in "Strategic Outlook" matters? This topic captures forward-looking discussions of growth, valuation, and investment opportunities. For example, reports sometimes pair a downgrade with language indicating attractive long-term prospects despite near-term headwinds. Consider this excerpt from a report with a positive future return prediction despite a downgrade: "Although the shares are an attractive long-term investment, we must caution that the severity of the current industry downturn increases the chance of further near-term earnings surprises." Linking the attribution to the broader evidence timing effect, the opportunity appears when analysts identify strong long-run fundamentals that are temporarily masked by negative short-term news. Supportively, when dividing the strategic discussion further into three dimensions of timeframe, sentiment, and focus, the Shapley value decomposition emphasizes the contribution of sentences related to positive long-term fundamentals.

A potential concern when using pre-trained LLMs for return prediction is lookahead bias, as their training data may inadvertently contain information about future events (Sarkar and Vafa, 2024). I contend that my results are robust to this concern for three reasons. First, LLMs are fundamentally designed to optimize language understanding rather than financial prediction capabilities. In other words, the pre-training process focuses on capturing semantic meaning through next-token prediction, not predicting stock returns. Second, if LLMs indeed learn and memorize future market information, we would expect to see the strongest predictability with the Company and Industry Overview content, as it provides the most comprehensive firm- and industry-specific information. Instead, this category generates insignificant alpha. Third, and more decisively, I directly test for lookahead bias by examining portfolio performance after each model's knowledge cutoff date with four different LLMs: BERT, RoBERTa, LLaMA2, and LLaMA3. The models are trained on data ending between December 2018 and March 2023. The earliest models (BERT and RoBERTa) provide over five years of true out-of-sample testing. If the predictability stemmed from information leakage, performance should deteriorate in these post-knowledge cutoff periods. Instead, the strategies generate even superior performances, yielding monthly long-short returns between 1.76% and 5.28% with Sharpe ratios ranging from 0.74 to 2.37.

This paper contributes to the literature on the value of information from sell-side analysts. Prior work mainly examines the cross-sectional predictability of analysts' numeric outputs—stock recommendations (Green, 2006; Christophe et al., 2010; Chen et al., 2023a), earnings forecasts (Lys and Sohn, 1990; Bordalo et al., 2019), target prices (Brav and Lehavy, 2003; Farago et al., 2023), and short-term trade ideas (Birru et al., 2022). I add to this line of research by showing that the narratives in analyst reports have significant investment value that is distinct from, and incremental to, these numeric forecasts. A related study, Lv (2024), uses a similar embedding-plus-Shapley design to study information content and value for strategic investors. In contrast, I focus on longer-horizon investment value that is not immediately incorporated into prices. The two papers are complementary: together they describe how markets process analyst research and how that processing creates value for investors.

Second, I contribute to the growing literature on text factors, which has largely extracted sentiment or topics from corporate disclosures, earnings-call transcripts, and news using dictionaries or topic models (e.g., Tetlock, 2007; Chen et al., 2022; Meursault et al., 2023; Bybee et al., 2023; Hirshleifer et al., forthcoming). I instead apply modern LLM embeddings to the nuanced, forward-looking narratives in sell-side reports. These representations capture richer semantics and context and can be decomposed at the topic level. The additive property allows me to isolate specific content categories, such as financial analysis, risk factors, and strategic outlook, and then quantify each category's standalone contribution to investment value. I validate the attribution by showing that a portfolio built solely on the highest-contributing topic content delivers the strongest univariate performance.

The rest of the paper proceeds as follows. Section 2.1 describes the data and methodology, including the prediction models, topic classification approach, and Shapley value decomposition framework. Section 3 presents the empirical results. Section 4 concludes.

2 Data and Methodology

2.1 Data

I assemble the sample from two primary sources. From Thomson Reuters Investext, I collect 1,194,330 sell-side analyst reports for S&P 1500 constituent firms over 2000–2023. Table 1 reports annual counts for report characteristics, brokerage coverage, and unique analysts. I then obtain analysts' numeric outputs—recommendations, EPS forecasts, target prices—and their announcement and revision histories from I/B/E/S.

To merge the datasets, I proceed in two steps. First, I map lead analysts in Investext to I/B/E/S analyst identifiers (AMASKCD) following Li et al. (2024). Second, I link each report to the corresponding I/B/E/S records within a five-trading-day window around the report date ([-2, +2]), following Lv (2024).

To benchmark against established predictors, I follow Gu et al. (2020) and construct 94 firm-level characteristics for NYSE, AMEX, and NASDAQ firms. The sample spans January 1957 to December 2023, and monthly returns are from CRSP. Table A2 defines the characteristics.

To identify the incremental signal relative to quantitative analyst information, I compile 18 factors from Chen and Zimmermann (2022) with significant investment values in analyst-produced information documented in the literature. These factors use analysts' forecasts, recommendations, or coverage changes to derive signals that predict stock performance. Table A3 provides a detailed description of the factors.

2.2 Methodology

2.2.1 Prediction Models

To capture narrative information, I quantify the perspectives embedded in analyst report text by mapping the report content into an embedding space. Compared with simple dimension tone measures in traditional NLP applications, the embeddings capture concise linguistic details and nuanced semantic meanings. Specifically, I input each report text into the LLaMA3-8B model and take the average of 32 transformer layer embeddings as text representation.

To extract value-relevant information, I use linear ridge regression to predict the next 12-month return following the report date. I deliberately use linear ridge regression models to emphasize the inherent information value of analyst reports rather than giving credit to sophisticated machine learning models. The ridge regression specifications are given by:

$$Ret_{it,12m} = \beta_0 + \beta' y^{\text{AI}}_{ijt} + \epsilon_{ijt}, \tag{1}$$

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left\{ \left\| Ret_{it,12m} - \beta_0 - y^{\text{AI}}_{ijt}\beta \right\|_2^2 + \theta \|\beta\|_2^2 \right\}, \tag{2}$$

where $Ret_{it,12m}$ is the next 12-month return of stock i following reports' release day t, and y^{AI}_{ijt} is the structured representation (4,096-dimension embeddings generated with the LLaMA3-8B model) of content from analyst report j on stock i released at day t.

Specifically, I use a monthly expanding window. For each out-of-sample month τ, I re-fit the ridge model on data available up to the prior month ($\tau - 1$), and select the penalty via five-fold time-blocked cross-validation over a log-spaced grid. The fitted model is then applied to reports released in month τ to produce out-of-sample 12-month return forecasts. The initial training sample consists of the first five years of data (2000–2004), and the out-of-sample forecasting period runs from January 2005 to December 2023.

2.2.2 Topic Modeling

I implement a two-step approach to interpretability. The first step assigns each sentence in the corpus to a semantically coherent, mutually exclusive topic; the second (described below) attributes predictive power to these topics.

Standard unsupervised methods, such as Latent Dirichlet Allocation (LDA) or k-means clustering of embeddings, are ill-suited for this task due to two primary challenges. First, the repetitive and formulaic language of analyst reports causes these models to identify stylistic patterns rather than distinct economic topics. Second, unconstrained clustering tends to generate industry-specific topics based on sectoral jargon, creating clusters that are not comparable across the full sample of reports.

To address both issues, I impose a fixed taxonomy following Lv (2024). The taxonomy specifies sixteen topics: *Executive Summary*, *Company Overview*, *Industry Analysis*, *Competitive Landscape*, *Income Statement Analysis*, *Balance Sheet Analysis*, *Cash Flow Analysis*, *Financial Ratios*, *Business Segments*, *Growth Strategies*, *Risk Factors*, *Management and Governance*, *ESG Factors*, *Valuation*, *Investment Thesis*, and *Appendices and Disclosures*. I add a residual category, *None of the Above*, to capture boilerplate, data tables, and stray sentences that do not map cleanly to any predefined topic, yielding seventeen labels in total.

I implement sentence-level classification with a fine-tuned BERT encoder, rather than a generative LLM, to exploit the stability and computational efficiency of supervised sequence classification at scale. The training set consists of 17,028 sentences manually labeled from 100 randomly selected reports. The trained model assigns each sentence a single topic label in $\{1, \ldots, 17\}$. On a stratified out-of-sample test set, the classifier achieves 89% accuracy. Residual misclassification should behave like classical measurement error and therefore attenuate, rather than inflate, subsequent estimates. I then apply the classifier to the full corpus, labeling 52,350,385 sentences and constructing report-level topic and meta-category profiles for the downstream return-prediction and attribution analyses.

I subsequently aggregate these fine-grained labels into five meta-categories that align with the main dimensions of sell-side research: company/industry overview, financial analysis, strategic outlook, risk and governance, and additional content. I aggregate the 17 fine-grained topics into five meta-categories for three reasons. First, interpretability: the meta-categories map cleanly to the economic functions of sell-side research. Second, statistical power and robustness: aggregation reduces multiple testing and attenuates classification noise (89% sentence-level accuracy implies some residual error that averages out within broader groups). Third, econometric and computational tractability: attribution based on groups rather than many narrow topics yields stable Shapley estimates and avoids overfitting from correlated sub-topics. Table A4 describes each meta-category and its constituent topics. Figure A1 visualizes salient terms.

2.2.3 Shapley Value Decomposition

The second stage attributes the investment value of the report signal to the seventeen topics defined in Table A4. Because topics may interact, an attribution that ignores joint effects is potentially misleading: two topics can be weak in isolation yet informative in combination. I therefore use a Shapley value decomposition (Lv, 2024), which averages marginal contributions over all permutations of topics and, by construction, satisfies efficiency (exact allocation of total value).

I construct topic representations at the sentence level to limit spillovers in the embeddings. For report j on firm i at time t, let y_{ijtd} denote the embedding of sentence d_{l} and let $\mathcal{D}_{p,ijt}$ be the set of sentences assigned to topic p by the classifier in Section 2.2.2. I form the token-weighted topic vector

$$\bar{y}^{\text{emb}}_{p,ijt} = \frac{\sum_{d\in\mathcal{D}_{p,ijt}} \text{Tokens}_d y_{ijtd}}{\sum_{d\in\mathcal{D}_{p,jit}} \text{Tokens}_d}, \quad p \in \{1,\dots,17\},$$

and then express the report-level embedding as a convex combination of topic blocks,

$$y^{\text{emb}}_{ijt} = \sum_{p=1}^{17} w_{p,ijt}\bar{y}^{\text{emb}}_{p,ijt}, \quad w_{p,ijt} = \frac{\sum_{d\in\mathcal{D}_{p,ijt}} \text{Tokens}_d}{\sum_{q=1}^{17}\sum_{d\in\mathcal{D}_{q,ijt}} \text{Tokens}_d}.$$

For any coalition $S \subseteq \{1,\dots,17\}$ of topics, I construct a representation that retains only the topics in S[2],

$$y^{\text{emb}}_{ijt}(S) = \sum_{p\in S} w_{p,ijt}\bar{y}^{\text{emb}}_{p,ijt}.$$

Holding coefficients fixed, I generate coalition-specific predicted returns,

$$\widehat{RET}^{(S)}_{ijt} = \hat{\beta}_0 + \hat{\beta}^\top y^{\text{emb}}_{ijt}(S),$$

and apply the same portfolio-formation rules as in the main tests (monthly sorts on $\widehat{RET}^{(S)}_{ijt}$, value-weighted deciles, long-short $H - L$ constructed and evaluated over the target horizon).

[2] I assign a zero block to the baseline feature value.

For each coalition I compute a performance functional, $V_t(S)$, such as the next-period $H - L$ return or the annualized Sharpe ratio of the strategy constructed from $\widehat{RET}^{(S)}$.

I fix the ridge coefficients estimated with full-topic estimates. This keeps the game fixed and lets the Shapley decomposition attribute the performance of one model to report content. I choose this setup for three reasons. First is cleaner interpretability. It isolates content contributions rather than gains from re-tuning when features are removed. The second reason is statistical stability. Re-fitting for every coalition would introduce new $\hat{\beta}^{(S)}$ and tuned penalties $\hat{\theta}^{(S)}$, greatly increasing variance and requiring nested time-series cross-validation over as many as $2^{17} = 131,072$ coalitions per month, which is computationally infeasible and would introduce significant estimation error. The third consideration is regularization artifacts. With ridge, dropping topics mechanically alters shrinkage on remaining coefficients, so attributions could reflect "regularization relief" rather than information in the dropped topic.

The Shapley contribution of topic p in month t is the average marginal improvement in the performance functional as topic p enters after any subset S that excludes it:

$$\phi_{p,t} = \sum_{S \subseteq \mathcal{P} \setminus \{p\}} \frac{|S|!(P - |S| - 1)!}{P!} \left[V_t(S \cup \{p\}) - V_t(S) \right], \quad P = 17, \quad \mathcal{P} = \{1, \ldots, 17\}.$$

For interpretability, I aggregate topic-level contributions to the five meta-categories introduced in Section 2.2.2. Let $\mathcal{M}_g$ collect the topic indices belonging to meta-category $g \in \{1, \ldots, 5\}$. By additivity,

$$\bar{\Phi}_g = \sum_{p \in \mathcal{M}_g} \bar{\phi}_p$$

gives the contribution of meta-category g. I compute its standard error by summing the corresponding monthly $\phi_{p,t}$ within g. This aggregation preserves exact additivity while yielding economically interpretable blocks.

3 Empirical Results

3.1 Quantifying Investment Value

This section quantifies the investment value of analyst reports. Section 3.1.1 examines return predictability, while Section 3.1.2 focuses on portfolio performance. Section 3.1.3 assesses the incremental value beyond established asset pricing factors, and Section 3.1.4 addresses potential lookahead bias concern.

3.1.1 Return Predictability

I begin by testing whether signals extracted from analyst reports can predict future stock returns. I construct these signals using large language models and machine-learning methods, and estimate the following report-level regression:

$$RET_{i,t+12} = \alpha_t + \beta' x_{i,t} + \varepsilon_{i,t+12}, \tag{3}$$

where $RET_{i,t+12}$ is the realized 12-month return for stock of stock i. $x_{i,t}$ denotes analyst information from each report, consisting of four measures: $\widehat{RET}_{12m}$ is the text-based 12-month return forecast from the ridge regression (1); REC_{rev} represents recommendation revision, calculated as the difference between the current report's recommendation and the analyst's last recommendation for the same stock in I/B/E/S; EF_{rev} denotes earnings forecast revision; and TP_{rev} represents target price revision. $Tone$ assesses report sentiment, using a fine-tuned BERT model that classifies each sentence as positive, negative, or neutral. Specifically, $Tone$ is calculated as:

$$Tone_i = \frac{N_i^+ - N_i^-}{N_i}, \tag{4}$$

where N_i^+ (N_i^-) represents the number of positive (negative) sentences in report i as classified by the BERT model, and N_i denotes the total number of sentences in the report. The specification includes year-month fixed effects, $\alpha_{m(t)}$, to absorb common time-series variation, ensuring that the coefficients capture cross-sectional predictive power. Standard errors are clustered by both firm and time to account for correlation in the residuals. The coefficient of interest, β_1, measures the predictive ability of the narrative signal after controlling for the analyst's explicit quantitative signals and report sentiment.

Table 2 reports the regression estimates. Panel A shows a pronounced, nearly linear relation between the text-based forecast $\widehat{RET}_{12m}$ and realized returns from one to twenty-four months. Each unit increase in $\widehat{RET}_{12m}$ raises the next-month return by 0.30 percent ($t = 1.85$). The effect grows to 0.60 percent for the twelve-month window ($t = 4.37$) and 1.14 percent for the twenty-four-month horizon ($t = 5.24$). This monotonic pattern suggests a gradual, rather than instantaneous, incorporation of the narrative's information into market prices.

Panel B contrasts the text-derived signal with standard numerical revisions. In univariate specifications (columns 1-4), none of the numerical outputs, like recommendation revisions, earnings forecast revisions, or target price revisions, nor a simple sentiment score, exhibits significant predictive power. By contrast, the narrative forecast $\widehat{RET}_{12m}$ is a strong predictor on its own (Column 5). Crucially, it retains its economic and statistical significance when all other signals are included as controls (Columns 6-7). In the fully specified model with year-month fixed effects, the coefficient on the narrative signal is 0.10 (t=3.34), and it remains highly significant even with the inclusion of more stringent fixed effects. In all multivariate specifications, the quantitative revisions and the Tone score are statistically and economically negligible.

Taken together, these findings indicate that analyst narratives contain valuable, long-horizon information that is priced slowly. Furthermore, the predictive power of the narrative-based signal is distinct from, and dominates, that of analysts' quantitative outputs.

To visualize the return dynamics following report issuance, I form portfolios based on daily sorts. Each day, all reports released are sorted into deciles based on their out-of-sample return predictions. I then track the subsequent performance of these portfolios by calculating their buy-and-hold abnormal returns (BHAR) using the characteristics-based benchmark approach of Daniel et al. (1997) (DGTW). The cumulative abnormal return (CAR) over a T-day horizon is defined as: Daniel et al. (1997):

$$CAR_{0,T} = \prod_{t=0}^{T} (1 + R_{it}) - \prod_{t=0}^{T} \left(1 + R_{it}^{DGTW}\right), \tag{5}$$

where R_{it} represents the raw return of stock i on trading day t, R_{it}^{DGTW} denotes the value-weighted return of stock i's benchmark portfolio on day t, and T indicates the end date of the accumulation period. $t = 0$ represents the first trading day when the market is open at or after a report is released.

Figure 1 plots the average cumulative abnormal returns for the top decile (most positive predictions) and bottom decile (most negative predictions) portfolios over the subsequent 252 trading days. The figure shows a clear and persistent divergence between the two portfolios. The top-decile portfolio earns a cumulative abnormal return of 3.75% over the year, while the bottom-decile portfolio underperforms its benchmark by -0.56%. The resulting long-short spread of 4.31% widens steadily over the period, providing strong visual evidence that the market is slow to incorporate the long-horizon information contained in the report narratives.

3.1.2 Performance of Long-Short Portfolios

I proceed with a portfolio analysis to evaluate the economic value of analyst report predictability. The trading strategy design incorporates two key elements. First, portfolios are rebalanced monthly to balance the tradeoff between capturing new information and minimizing transaction costs. Second, motivated by the persistence in return predictability documented in Panel A of Table 2, I aggregate analyst report predictions over extended historical windows to capture sustained signals in analyst reports.

Specifically, at the beginning of each month, I construct portfolios with the following procedure: First, for each stock, I calculate the average predicted return from all analyst reports issued during the past LB months, where $LB \in \{9, 12, 18, 24\}$. I then sort stocks into decile portfolios based on these averaged predictions. Within each decile portfolio, stocks are weighted by their previous month-end market capitalization to mitigate the influence of microcaps. I then construct a zero-cost long-short portfolio that takes a long position in the highest predicted return decile (decile 10) and a short position in the lowest predicted return decile (decile 1).

Table 3 presents performance statistics of value-weighted decile portfolios across lookback periods (LB) of 9, 12, 18, and 24 months. Across all LB, the decile spread is economically large and statistically significant: the $H - L$ portfolio earns $0.87\% - 1.16\%$ per month with Sharpe ratios of $0.62 - 0.69$. Risk adjustment using the Fama-French six factors yields positive, significant alphas of $0.52\% - 0.79\%$ per month (t-statistics between 1.99 and 2.64). The cross-decile pattern of alphas indicates that performance is primarily driven by the long leg rather than the short leg, consistent with the asymmetric price paths in Figure 1. Based on this evidence, I use the 12-month lookback window as the benchmark in subsequent analyses: it delivers the highest $H - L$ alpha t-statistic (2.64) and aligns with the model's 12-month return-prediction horizon.

Figure 2 plots cumulative returns for the benchmark $LB = 12$ strategy from January 2005 to June 2024. The self-financing long–short portfolio compounds to 846% over the sample, compared with a 399% cumulative excess return on the value-weighted market. Decomposing the legs, the long side compounds to 2,274% over the period, while the short side delivers 151% Thus, the investment value stems from picking stocks with strong future prospects, which is consistent with the portfolio statistics in Table 3. The strategy experiences significant drawdowns during the global financial crisis and again post-2022, conditions in which fundamental signals can be temporarily overshadowed by systemic risk aversion and forced liquidations.

3.1.3 Incremental Investment Value

A key question is whether the signals from analyst reports provide value beyond exposure to established market factors. While analysts are experts, their reports could simply reflect common factor risks like value or momentum. To test for unique, idiosyncratic information, I compute riskadjusted returns (α) for my strategy against four prominent factor models: the Fama and French (2015) five factor model (α_{F5}), the Fama and French (2018) six factor model (α_{F6}), the q -factor model of Hou et al. (2015) (α_{HXZ}), and the behavioral factor model of Daniel et al. (2020) (α_{DHS}). I then conduct a broader analysis to see if the report signals contain information beyond the 94 factors documented in Gu et al. (2020).

I also examine whether the narrative-based signals are informative relative to traditional quantitative signals produced by analysts. To do this, I test for incremental value over 18 established analyst-based factors, such as EPS forecast revisions and recommendation changes, from the Chen and Zimmermann (2022) database.

In Table 4, I evaluate the predictive power of analyst reports relative to and in combination with established return predictors. The analysis examines three main signal families: analyst report full content predictions (RP), the average return series of 18 analyst-based factors (ANA), and predictions from a 4-layer artificial neural network (ANOM), the best-performing machine learning model in Gu et al. (2020). For each strategy and combination, I report mean returns, Sharpe ratio (SR), and information ratio (IR), with corresponding t-statistics in parentheses. The Information Ratio of the portfolio i relative to a benchmark b is calculated as:

$$IR_{i,b} = \frac{\bar{r}_{i,b}}{\sigma_{i,b}}, \tag{6}$$

where $\bar{r}_{i,b}$ is the average active return (difference between portfolio i and benchmark b returns) and $\sigma_{i,b}$ is the standard deviation of these active returns, also known as tracking error. In Table 4, I compute the IR of combined strategies relative to individual components to measure the incremental value of information sources. For example, the IR of "RP + ANA" measures the incremental gain from combining report predictions with analyst-based factors, relative to using "ANA" signals only.

Panel A first establishes the standalone performance of each information source. Report text generates significant risk-adjusted returns across various factor models, with alphas ranging from 0.73 to 1.21. The numerical analyst-based factors (ANA) show modest but significant predictive power (alphas between 0.16 and 0.28). The machine learning-based anomaly portfolio (ANOM) exhibits strong performance with alphas between 1.15 and 1.40, suggesting effective capture of fundamental signals.

Panel B addresses the central question of incremental value. When combining report predictions with analyst-based factors (RP + ANA), the portfolio maintains significant alphas (0.50-0.68) and shows meaningful incremental gain with an IR of 0.73. Most notably, incorporating both analyst and fundamental factors (RP + ANA + ANOM) produces the highest Sharpe ratio (1.60) and consistent alphas across factor models (0.77-0.92), with an IR of 1.23 (t=4.05).

These results suggest that analyst reports contain statistically and economically significant incremental information not captured by traditional analyst outputs like forecast revisions or by machine learning models trained on fundamental data. The strong incremental performance and information ratios indicate that textual information in research reports offers a distinct and complementary source of investment value.

3.1.4 Lookahead Bias

A key concern in demonstrating out-of-sample profitability from analyst reports is lookahead bias. Because LLMs are pre-trained on massive text corpora, there is a risk that they could have indirectly learned from future stock performance data.

To address this concern directly, I conduct a stringent analysis using only post-knowledge-cutoff samples. This test employs four distinct LLMs of varying scale and architecture: the pioneering BERT (110M parameters) and RoBERTa (355M parameters) models, the larger LLaMA2-13B, and the more recent LLaMA3-8B. For each model, I form portfolios and evaluate their performance only in the period after the model's training data cutoff date. This creates a true out-of-sample test free from potential contamination. The approximate cutoff dates for these models are well before the end of my sample period, providing a multi-year window for a clean out-of-sample test.

Table 5 presents the results. Strikingly, the performance in this true out-of-sample period is even stronger than in the full sample. Examining the high-minus-low (H-L) portfolio returns, all models generate substantial positive returns ranging from 1.59% (LLaMA2) to 2.70% (LLaMA3) with corresponding Sharpe ratios between 0.74 and 1.28. Under the leakage hypothesis, performance should deteriorate once the evaluation moves beyond each model's training horizon (especially for models with earlier cutoffs). I find the opposite. The post-cutoff tests indicate that profitability reflects information extracted from contemporaneous analyst narratives, not memorized future outcomes from pretraining text.

3.2 Where Does the Investment Value Come From?

I next ask what types of analysts and firms receive high text-based return forecasts. Section 3.2.1 characterizes the cross-section of signals. Section 3.2.2 attributes the signal to specific report content. Section 3.2.3 studies state dependence by conditioning on contemporaneous news.

3.2.1 Which stock is picked?

Table 6 presents a cross-sectional analysis, comparing the characteristics of firms and analysts in the top and bottom deciles of the narrative-based return forecast $\widehat{RET}_{12m}$. The table reports means and medians for each group, along with t -tests and Wilcoxon rank-sum tests for differences.

Panel A indicates limited dispersion in analyst attributes. While reports with high forecasts are issued by analysts with slightly narrower coverage (fewer industries and firms) from smaller brokerages, the economic magnitudes of these differences are small, even when statistically detectable. Other key attributes, such as analyst experience and affiliation with a "Top 10" brokerage, show no meaningful difference across the groups. This pattern suggests the narrative signal is not merely a proxy for analyst status or coverage breadth.

Panel B shows a mild tilt in firm characteristics. Firms associated with high text-based forecasts are modestly tilted toward profitable growth. Relative to the Low group, High-signal firms are larger (mean log market cap 16.18 vs. 16.06), have lower book-to-market ratios (0.37 vs. 0.45), and exhibit higher gross profitability (0.39 vs. 0.35). Investment is similar across groups (1.01 vs. 0.98), while idiosyncratic volatility is modestly higher for the High group (0.02 vs. 0.01). These differences are statistically significant by both t-tests and Wilcoxon rank-sum tests (see p-values in the table), but their magnitudes are economically small.

To benchmark these patterns against standard risk factors, I regress monthly portfolio returns formed on $\widehat{RET}_{12m}$ on the Fama-French six factors (market, size, value, profitability, investment, momentum). Table 7 shows that the High-Low portfolio is close to market-neutral but not flat (market beta ≈ 0.28). Size exposure is near zero. Loadings on value and investment are strongly negative, consistent with a growth and aggressive-investment tilt. The profitability loading is positive but modest, and momentum is statistically weak. The long-only leg exhibits the corresponding pattern: a high market beta (≈ 1.20), negligible size exposure, pronounced negative HML and CMA loadings, and only a small profitability tilt with little relation to momentum.

Taken together, the cross-sectional comparisons and factor regressions imply that the text-based forecast primarily selects profitable growth firms with relatively aggressive investment, rather than small-cap or momentum. The exposures are moderate in size, which, together with the weak role for momentum, suggests that the predictive content of the narrative signal is not simply a repackaging of standard factor bets. Instead, it appears to capture incremental information about longer-horizon fundamentals that is only partially aligned with conventional characteristics.

The observed correlations between the text-based return forecast and firm characteristics could be incidental, or it could indicate that the signal's predictive power is concentrated in certain types of firms. To distinguish between these possibilities, I analyze how the forecast's performance varies across subsamples sorted on firm attributes. Figure 3 plots

DGTW-adjusted buy-and-hold abnormal returns over a 252-trading-day window for double-sorted portfolios. In each panel, the sample is first partitioned by the monthly median of a characteristic, and then within each group, I trace the average abnormal returns for portfolios formed on the top and bottom deciles of the narrative signal.

I first examine the return predictability of analyst reports across four firm characteristics: 1) large vs small stocks; 2) high versus low book-to-market ratio; 3) high versus low idiosyncratic volatility; 4) high versus low gross profitability. Panel (a) shows that reports on large-capitalization stocks generate significantly higher abnormal return spreads compared to small-cap stocks (4.14% vs 1.74%), with the difference particularly pronounced for stocks predicted to have high returns. This pattern suggests that despite the greater analyst coverage and presumed market efficiency in the large-cap segment, reports from skilled analysts still contain meaningful investment opportunities among these stocks. The valuation results in Panel (b) indicate that high valuation firms offer greater potential for profitable analyst recommendations compared to low valuation firms. Panel (c) shows that the return spread is markedly larger among high-idiosyncratic-volatility firms. The high-prediction portfolio appreciates steadily while the low-prediction portfolio drifts flat to negative, producing a wide, monotonic gap; the corresponding spread within the low-volatility stratum is materially smaller. Panel (d) stratifies by gross profitability and indicates that predictability is concentrated in profitable firms. High-profitability stocks with favorable predictions experience a strong and persistent post-report drift, whereas the spread among low-profitability firms is muted and slower to emerge. Taken together, the four firm-side panels imply that the text-based signal is most effective for large-cap, growth, high-volatility, and high-profitability firms.

Conditioning on firm characteristics, Figure 3 shows that the post-publication DGTW-adjusted buy-and-hold abnormal return spreads are concentrated in specific strata. Along firm size (panel a), predictability is stronger for large-capitalization firms: the +252-day H-L spread is 4.14% for large caps versus 1.74% for small caps, with the differential largely driven by the high-prediction leg. Along valuation (panel b), splitting on book-to-market indicates larger spreads for growth (low BM ratio) firms than for value (high BM ratio) firms, consistent with the cross-sectional tilt toward lower BM ratios in Table 6 and the negative HML exposure in factor regressions. Along with idiosyncratic volatility (panel c), the spread is markedly larger in the high-volatility stratum: the high-prediction portfolio appreciates steadily while the low-prediction portfolio is flat to negative, yielding a wide, monotonic gap. The low-volatility subsample exhibits a materially smaller divergence. Along profitability (panel d), predictability concentrates in high-profitability firms: stocks with favorable forecasts and high gross profitability display a strong and persistent post-report drift, whereas the corresponding spread within the low-profitability stratum is muted and slower to emerge. Overall, the conditional evidence indicates that the text-based signal is most effective for large-cap, growth, high-volatility, and high-profitability firms, consistent with the portfolio loadings on value and investment reported above.

In addition, I also test the predictability in the cross-section of analyst skills and experience. Analyst skill, measured by historical EPS forecast accuracy and shown in Panel e, appears to be a crucial determinant of report value. High-skill analysts generate significantly larger abnormal return spreads (5.74%), particularly for their most optimistic recommendations, while reports from low-skill analysts show significantly smaller investment value in terms of return spreads (4.44%). Finally, Panel f demonstrates that analyst experience plays a vital role in report effectiveness. Experienced analysts' reports, especially those predicting high returns, generate substantial abnormal returns approaching 6.14% over the holding period. In contrast, reports from less experienced analysts show less ability to identify profitable investment opportunities.

In summary, the cross-sectional analyses reveal a clear picture of where the narrative's investment value originates. The signal is most potent when applied to large, profitable growth firms with high idiosyncratic risk. Furthermore, this effect is significantly amplified when the narrative is authored by analysts who are more skilled and experienced.

3.2.2 What content matters?

I now investigate the sources of investment value in analyst reports by decomposing portfolio returns and Sharpe ratios across five topic categories, as described in Section 2.2.2. Using yearly fitted ridge models, I calculate each topic's contribution to portfolio performance by measuring the average decrease in performance across all possible topic combinations. I group together the topic Shapley values with specific categories and report the sum of their constituent importance.

Table 8 presents both the distribution of content and its contribution to portfolio performance. Financial analysis and company/industry overview dominate the content, comprising 36.56% and 28.53% of total coverage, respectively. Strategic outlook and risk and governance form a second tier of coverage at 15.13% and 14.14% of sentences, while additional content accounts for the remaining 5.63%.

The Shapley value decomposition reveals a different pattern. Strategic Outlook emerges as the most valuable category, accounting for 41.34% and 31.43% of the portfolio's Sharpe ratio and returns, respectively, despite not being the most extensively discussed component. Company and Industry Overview maintains proportional importance, contributing 27.61% to the Sharpe ratio and 26.92% to returns. Risk and Governance shows outsized impact on returns relative to its discussion frequency, accounting for 11.21% of Sharpe ratio value and 21.36% of returns. Additional Content has minimal impact on performance.

Perhaps the most surprising result is the low contribution of Financial Analysis, given findings in Lv (2024) that emphasize high information content in income statement analysis. Despite constituting the most extensively covered category (36.56% of content), Financial analysis discussions contribute only to 16.39% of the Sharpe ratio and 19.53% of returns. One reason may be that financial data are more rapidly priced in, leaving less room for drift or mispricing to persist. In contrast, strategic and forward-looking insights may only gradually be reflected in valuations, thus creating greater opportunity for abnormal returns.

The findings in Table 8 suggest that the investment value of analyst reports stems primarily from their strategic insights and industry analysis rather than traditional financial analysis. The substantial contribution of strategic content to portfolio performance, despite its relatively lower coverage, indicates that analysts' forward-looking assessments and strategic perspectives may be particularly valuable for investment decisions.

Table 9 presents a detailed analysis of strategic outlook content along three key dimensions - timeframe, sentiment, and analytical focus - using Shapley value decomposition. For each dimension, I classify sentences using a structured prompt that categorizes timeframe (long-term, short-term, or both), sentiment (negative, neutral, or positive), and focus (risk versus fundamental analysis) with the following prompt:

Prompt:
Please read the following sentence from a sell-side analyst report (investment thesis section) carefully and classify it into three numeric labels:
Timeframe: 1 = Long-term (multi-quarter or multi-year outlook); 2 = Short-term (next quarter or near-term); 3 = Both (if it clearly addresses both short- and long-term).
Sentiment: 1 = Positive potential; 0 = Neutral; 1 = Negative/Bearish.
Focus: 1 = Fundamentals (e.g., belief in growth strategy, industry tailwinds, earnings); 0 = Cautious risk (e.g., warnings, near-term headwinds, legal/regulatory risk).
Important note: If the sentence mentions both short-term risk and a fundamental (longer-term) driver, prioritize the fundamental aspect and mark the third label as **1** (Fundamentals).
Output format: Only output the three numbers in the format: `[X, Y, Z]`.
Here is the sentence: {sentence}

The results reveal several notable patterns. First, long-term oriented discussions dominate the investment value, accounting for approximately half of both the Sharpe ratio (49.51%) and return (50.20%) contributions. Short-term discussions contribute about a quarter of the predictive power, while discussions spanning both timeframes account for the remaining portion.

Second, the sentiment analysis shows that positive outlooks generate the largest share of investment value, contributing 49.94% to the Sharpe ratio and 50.86% to returns. Neutral and negative sentiments have roughly equal contributions, each accounting for about a quarter of the portfolio performance metrics.

Perhaps most strikingly, the focus decomposition shows that fundamental analysis overwhelmingly drives investment value, accounting for 87.00% of the Sharpe ratio and 87.75% of return contributions. Risk-focused discussions, while important for comprehensive analysis, contribute relatively little to portfolio performance metrics.

These findings suggest that the predictive power of strategic outlook sections stems primarily from long-term, positive discussions focused on fundamental analysis, rather than from risk warnings or short-term forecasts. This pattern aligns with the notion that durable competitive advantages and long-term growth opportunities require more analytical skill to identify and value correctly than short-term fluctuations or risk factors.

A follow-up question would be, since Shapley value attribution pins down the content contributing the most to investment value, can we separate this most valuable information and form a portfolio free of text noise? In Table 10, I report the value-weighted decile portfolio performance sorted by category-specific content. I focus on four categories, excluding additional content that primarily consists of latent sentences.

Strategic Outlook content generates the strongest portfolio performance, with the long-short strategy delivering a mean return of 1.41% per month and a Sharpe ratio of 0.93. The outperformance appears to be driven by both the long and short legs, with the highest decile earning 1.87% monthly and showing significant alpha (0.72, t = 2.85). This superior performance aligns with the earlier Shapley value decomposition, which identifies Strategic Outlook as the most valuable category.

The remaining categories, Company and Industry Overview, Financial Analysis, and Risk and Governance, generate more modest performance metrics. Long-short returns range from 0.52% to 0.58% monthly and Sharpe ratios are between 0.37 and 0.41. While the highest deciles of these categories show some predictive power, particularly Risk and Governance with an alpha of 0.32 (t = 1.97), their overall performance is substantially weaker than Strategic Outlook. This pattern is again consistent with the Shapley value decomposition. It suggests that traditional financial analysis, industry insights, and risk assessments are less informative about future returns than forward-looking strategic insights.

I next isolate the incremental contribution of Strategic Outlook. The Shapley attribution identifies it as the dominant content block, and the univariate strategy that uses only

Strategic Outlook predictions validates this allocation. As reported in Table A5, the Strategic Outlook portfolio delivers information ratios of 0.96–1.41 when evaluated against both the 94 fundamental predictors of Gu et al. (2020) and the 18 analyst-based factors of Chen and Zimmermann (2022). Thus, focusing on the single most value-relevant component produces more efficient portfolios than full-text signals, consistent with a concentrated rather than diffuse source of information.

Table A6 shows that this result is not driven by lookahead bias. Across alternative language-model representations, the Strategic Outlook strategy attains a post-knowledge cutoff Sharpe between 1.26 and 2.37. Taken together, the two tables reinforce the central interpretation: long-horizon investment value in analyst narratives is concentrated in forward-looking strategic assessments, and that content alone is sufficient to generate robust, high-efficiency portfolios.

3.2.3 State-Dependent Predictability: Good vs. Bad News

Having established the characteristics of the signal, I next examine the conditions under which its predictability arises. Since analyst reports are often issued around major corporate announcements, a key question is whether the narrative signal's performance is simply another manifestation of a known anomaly, such as post-earnings-announcement drift (PEAD). If so, the signal would reflect an underreaction to public earnings news rather than novel, long-horizon information contained in the narrative itself.

To separate these channels, I implement a double-sorting analysis based on earnings surprises and return predictions, presented in Figure 4. First, reports are classified into 'SUE(Small)' or 'SUE(Large)' groups based on whether their stocks' most recent earnings surprise falls below or above the monthly sample median. Following Livnat and Mendenhall (2006), earnings surprise (SUE) is calculated as the actual EPS minus the last consensus EPS forecast before the earnings announcement, with higher values indicating positive news. Second, within each SUE group, I sort reports into deciles based on their predicted returns. Recall that the return predictability is driven by long leg (positive predictions). If the observed predictability were attributable to a price-drift like PEAD, then reports with high SUE (i.e., good earnings news) should exhibit a greater spread between the 'High' and 'Low' groups. Because earlier evidence shows that predictability is driven mainly by the long leg, the PEAD hypothesis implies a larger High–Low spread in the high-SUE group.

Contrary to the prediction of the PEAD hypothesis, the results in Panel (a) show that the narrative signal's predictive power is significantly stronger following adverse earnings news. The cumulative long-short spread over the subsequent year is 7.52% for reports issued after low earnings surprises, compared to 4.30% for those issued after high surprises. This finding is inconsistent with a simple price drift explanation and instead suggests that the narrative contains information orthogonal to the short-term earnings announcement.

This pattern of heightened predictability after adverse signals extends to the analysts' own quantitative revisions. I test this by conditioning on three different numerical signals: recommendation changes (Panel b), EPS forecast revisions (Panel c), and target price revisions (Panel d). In all three cases, the predictive spread of the narrative-based signal is substantially larger when the contemporaneous quantitative signal is unfavorable (e.g., following a downgrade versus an upgrade, or a below-median target price revision). As in prior analyses, this outperformance is driven primarily by the long leg of the strategy and emerges gradually over the subsequent year.

Taken together, this conditional analysis provides strong evidence for state-dependent informativeness. The narrative content of analyst reports is most predictive, and thus most valuable, precisely when contemporaneous public news and the analyst's own quantitative signals are negative. This pattern is inconsistent with a mechanical continuation of good news. Rather, it supports the view that analysts' narratives provide forward-looking assessments that the market is particularly slow to incorporate during periods of negative sentiment.

Motivated by the state-dependence results, I test whether the documented predictability is merely a sophisticated form of sentiment, or perhaps a reversal of an initial announcement-window overreaction. Each month, I form value-weighted decile portfolios by sorting stocks on four sentiment-proximate signals, averaged over the prior twelve months of reports for each stock: (i) the two-day DGTW-adjusted abnormal return around report dates, $CAR_{[0,+1]}$; (ii) recommendation revisions; and (iii) headline tone and (iv) body tone from a fine-tuned BERT classifier. I then evaluate next-month performance, including mean excess returns, volatility, Sharpe ratios, and alphas relative to the Fama and French (2018) six factors.

The results, presented in Table 11, do not support a sentiment-based or overreaction-based explanation. A strategy based on short-term announcement returns $\left(CAR_{[0,+1]}\right)$ fails to generate a premium. The long-short alpha is negative and statistically insignificant (-0.36% per month, t $= -1.6$). Likewise, strategies based on document-level sentiment, either headline or body tone, show no monotonic relationship between tone and returns, and their long-short alphas are economically and statistically negligible. While a strategy based on recommendation revisions shows slightly more promise in raw returns, its risk-adjusted alpha remains weak (0.24% per month, t $= 1.6$). In all four cases, these sentiment-proximate strategies fail to deliver compelling risk-adjusted performance.

The evidence strongly suggests the narrative signal derives its value from capturing complex, forward-looking assessments of firm fundamentals, not from short-term market sentiment. A strategy based on announcement-window returns generates no alpha, ruling out a simple overreaction story. Similarly, strategies based on coarse document-level tone are unprofitable, indicating that a simple positive/negative sentiment dimension misses the crucial information contained in the full narrative.

4 Conclusion

I document novel investment value embedded in the textual narratives of analyst reports. By representing these narratives as high-dimensional LLM embeddings, I show they predict future stock returns. A portfolio strategy based on this signal delivers economically and statistically significant alpha that is incremental to analysts' quantitative outputs and a vast set of established factor- and characteristic-based predictors.

The signal's predictive power is not uniform. It is highly state-dependent and concentrated in specific cross-sections. Predictability is stronger precisely when contemporaneous news is adverse. For example, following negative earnings surprises or recommendation downgrades. The effect is also amplified for reports authored by more skilled and experienced analysts and is most pronounced for large, profitable growth firms. Factor regressions show moderate tilts (negative value and investment), yet the alpha remains after standard adjustments. Coarse report sentiment measures and announcement-window reactions do not reproduce the effect.

Using a Shapley value decomposition built from sentence-level topics aggregated to five meta-categories, I attribute performance to economically interpretable content. Strategic outlook dominates, with the largest contributions from investment thesis and valuation language that emphasizes longer-run fundamentals. A strategy based solely on this Strategic Outlook content exceeds the full model's alpha, indicating that the incremental information is highly concentrated in this forward-looking section.

These findings provide direct evidence that analyst narratives embed forward-looking assessments that the market incorporates gradually. Methodologically, the paper shows how modern language models can recover and attribute economic content from unstructured text without sacrificing interpretability. Substantively, they show that generative AI complements rather than replaces human expertise. By scaling and systematizing the judgments of financial analysts, these tools allow the rich information embedded in qualitative research to be measured, tested, and deployed in a disciplined way.

References

Birru, Justin, Sinan Gokkaya, Xi Liu, and René M Stulz, 2022, Are analyst short-term trade ideas valuable?, *The Journal of Finance* 77, 1829–1875.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, 2019, Diagnostic expectations and stock returns, *The Journal of Finance* 74, 2839–2874.

Brav, Alon, and Reuven Lehavy, 2003, An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics, *The Journal of Finance* 58, 1933–1967.

Bybee, Leland, Bryan Kelly, and Yinan Su, 2023, Narrative asset pricing: Interpretable systematic risk factors from news text, *The Review of Financial Studies* 36, 4759–4787.

Chen, Andrew Y., and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 27, 207–264.

Chen, Hailiang, Byoung-Hyoun Hwang, and Zhuozhen Peng, 2023a, Revisiting the cross-section of expected stock returns: Evidence from a textual analysis of buy recommendations, *Nanyang Business School Research Paper* .

Chen, Luyang, Markus Pelger, and Jason Zhu, 2023b, Deep learning in asset pricing, *Management Science* .

Chen, Yifei, Bryan T Kelly, and Dacheng Xiu, 2022, Expected returns and large language models, *Available at SSRN 4416687* .

Christophe, Stephen E, Michael G Ferri, and Jim Hsieh, 2010, Informed trading before analyst downgrades: Evidence from short sellers, *Journal of Financial Economics* 95, 85–106.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of Finance* 52, 1035–1058.

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short-and long-horizon behavioral factors, *The Review of Financial Studies* 33, 1673–1736.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and Kenneth R French, 2018, Choosing factors, *Journal of financial economics* 128, 234–252.

Farago, Adam, Erik Hjalmarsson, and Ming Zeng, 2023, Analysts are good at ranking stocks, *Available at SSRN 4495163* .

Green, T Clifton, 2006, The value of client access to analyst recommendations, *Journal of Financial and Quantitative Analysis* 41, 1–24.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Hirshleifer, David, Dat Mai, and Kuntara Pukthuanthong, forthcoming, War discourse and the cross section of expected stock returns, Technical report.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

Jegadeesh, Narasimhan, Joonghyuk Kim, Susan D Krische, and Charles MC Lee, 2004, Analyzing the analysts: When do recommendations add value?, *The Journal of Finance* 59, 1083–1124.

Jha, Manish, Hongyi Liu, and Asaf Manela, 2020, Does finance benefit society? a language embedding approach, *A Language Embedding Approach (July 18, 2020)* .

Li, Kai, Feng Mai, Rui Shen, Chelsea Yang, and Tengfei Zhang, 2024, Dissecting corporate culture using generative ai–insights from analyst reports, *Available at SSRN 4558295* .

Livnat, Joshua, and Richard R Mendenhall, 2006, Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts, *Journal of Accounting Research* 44, 177–205.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of Finance* 66, 35–65.

Lv, Linying, 2024, The value of information from sell-side analysts, *arXiv preprint arXiv:2411.13813* .

Lys, Thomas, and Sungkyu Sohn, 1990, The association between revisions of financial analysts' earnings forecasts and security-price changes, *Journal of Accounting and Economics* 13, 341–363.

Meursault, Vitaly, Pierre Jinghong Liang, Bryan R Routledge, and Madeline Marco Scanlon, 2023, Pead. txt: Post-earnings-announcement drift using text, *Journal of Financial and Quantitative Analysis* 58, 2299–2326.

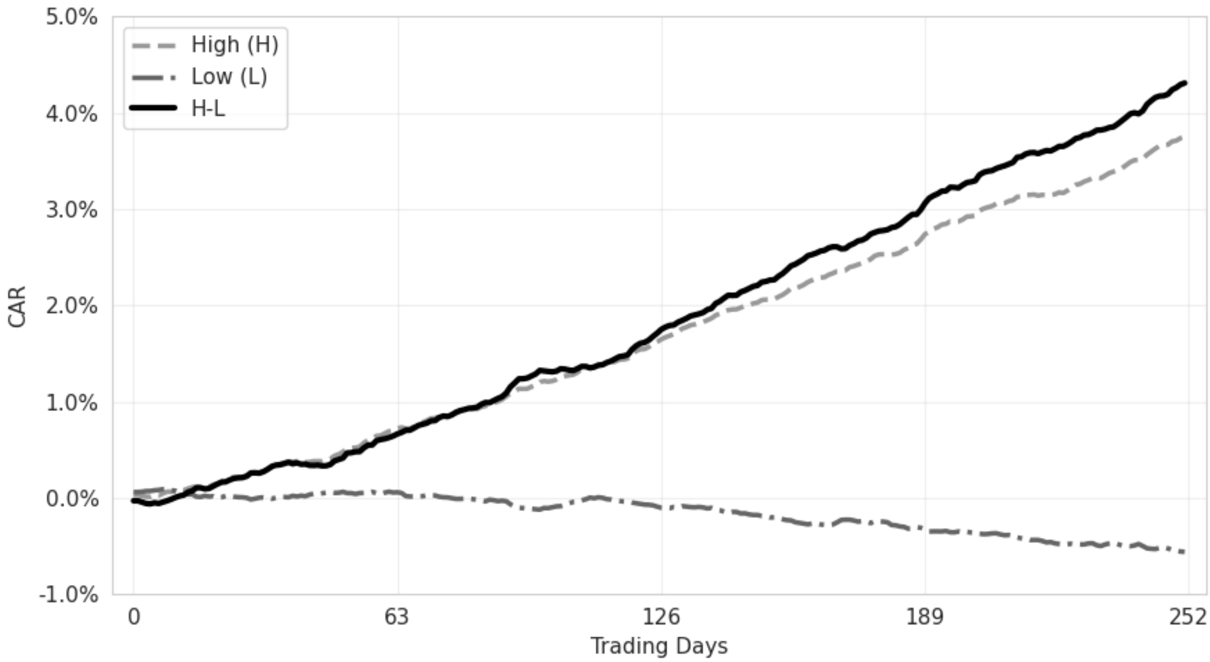
Sarkar, Suproteem K, and Keyon Vafa, 2024, Lookahead bias in pretrained language models, *Available at SSRN* .

Siano, Federico, 2025, The news in earnings announcement disclosures: Capturing word context using llm methods, *Management Science* .

Tetlock, Paul C, 2007, Giving content to investor sentiment: The role of media in the stock market, *The Journal of Finance* 62, 1139–1168.

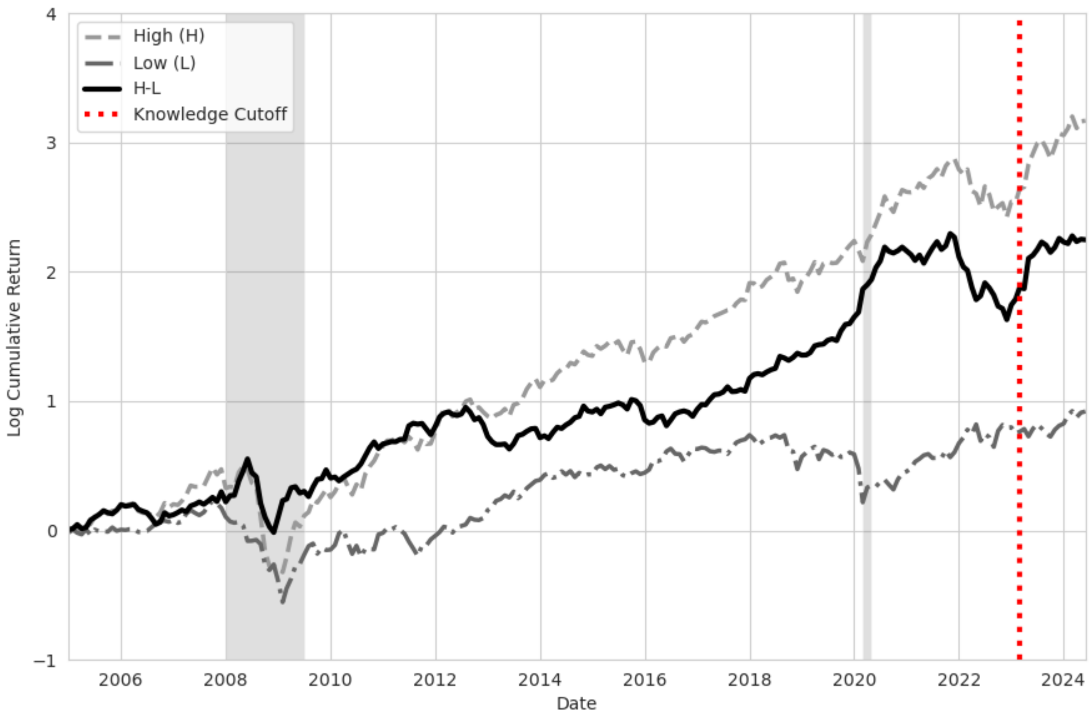
Womack, Kent L, 1996, Do brokerage analysts' recommendations have investment value?, *The Journal of Finance* 51, 137–167.

Figure 1: Report-based Forecasts and Abnormal Stock Returns



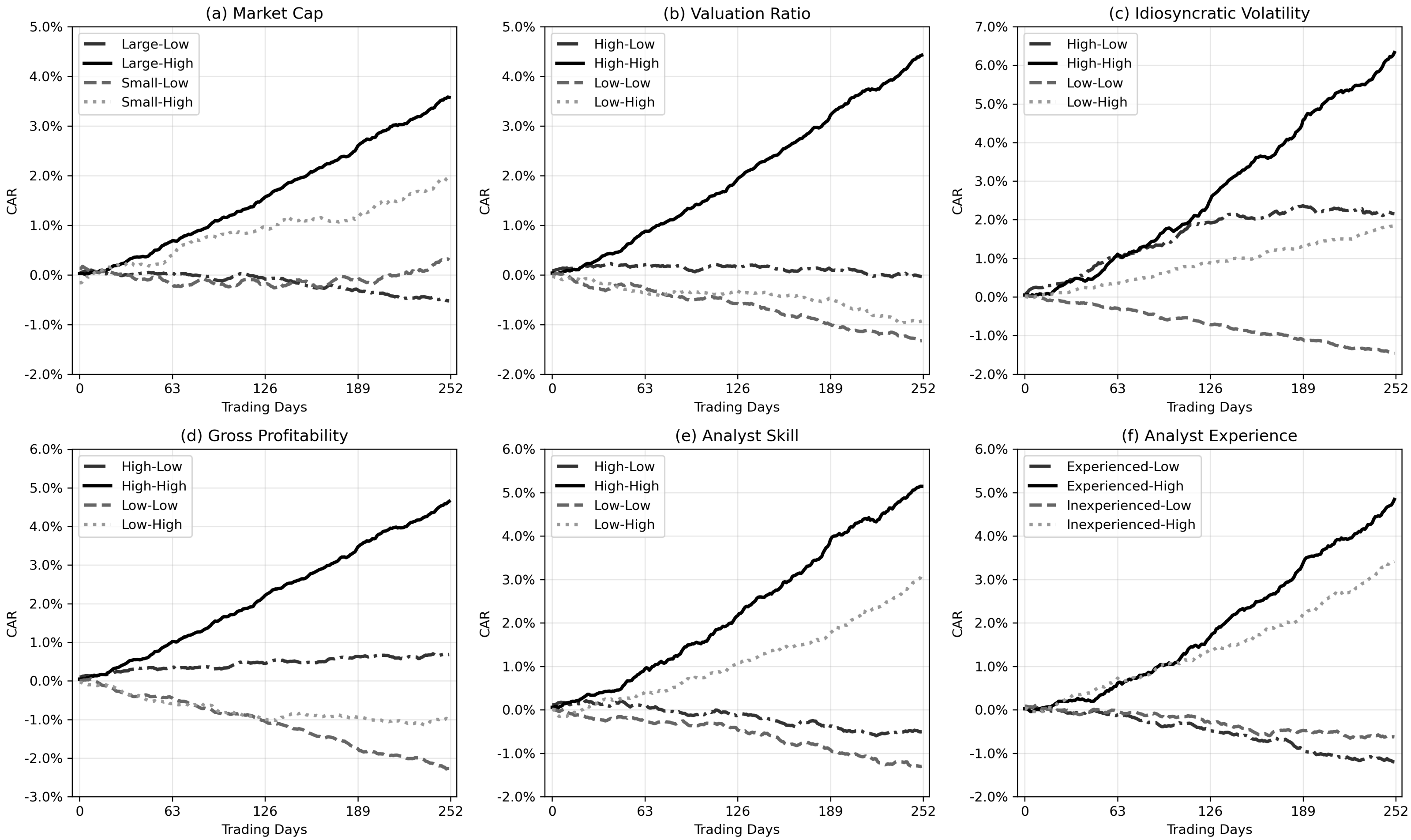
Note: This figure presents DGTW characteristic-adjusted buy-and-hold abnormal returns following research report releases, measured over a [0, +252] trading day window after each announcement. Reports are sorted into deciles based on their predicted 12-month returns. The 'High (H)' group consists of reports predicting returns in the highest decile within each trading day, while the 'Low (L)' group comprises reports predicting returns in the lowest decile. The lines represent value-weighted abnormal returns of the stocks corresponding to these reports, and 'H-L' denotes the return spread between the High and Low groups. The sample spans from 2005 to 2023.

Figure 2: Cumulative Log-return of Portfolios Sorted on Report Forecasts



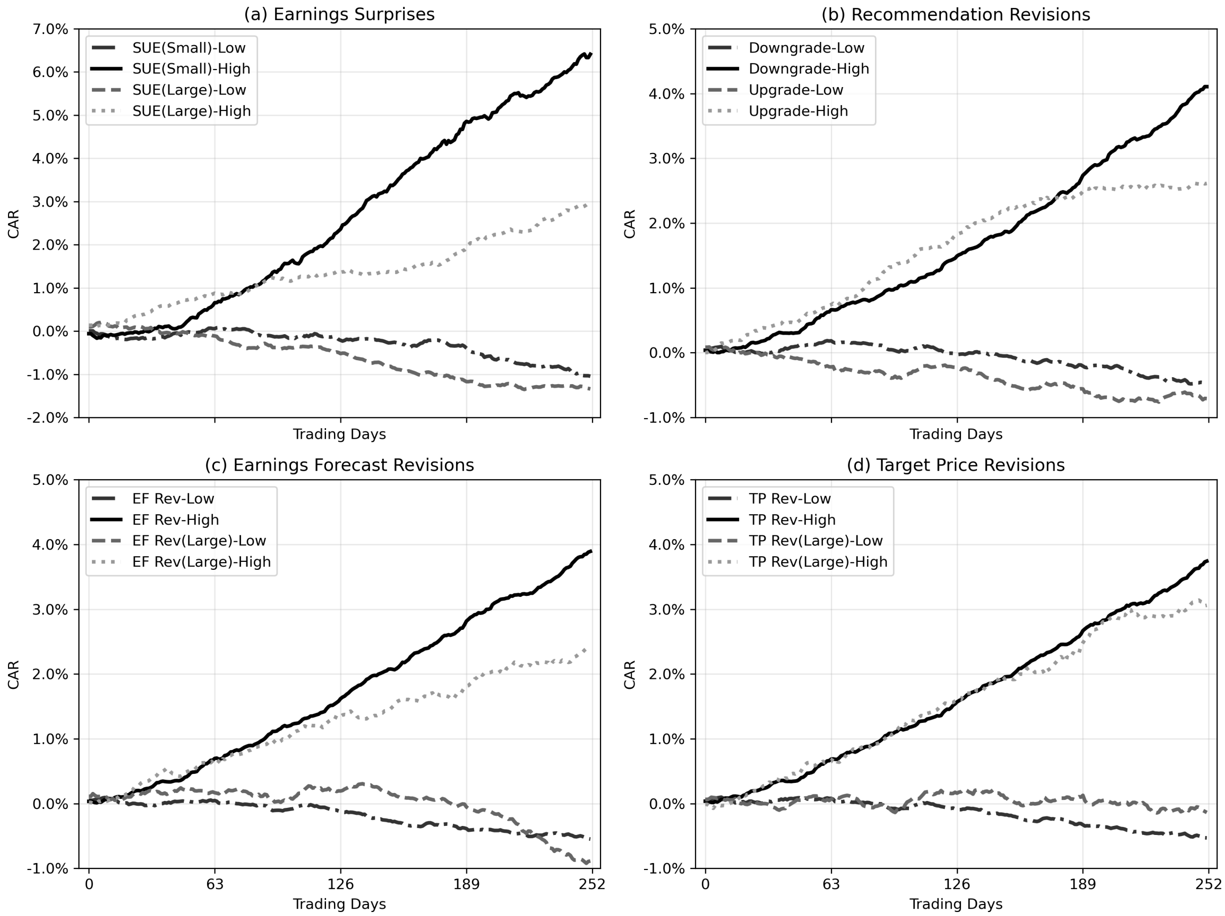
Note: This figure shows cumulative log excess returns for portfolios constructed based on analyst report forecasts from the previous 12 months. For each panel, three portfolios are presented: the highest decile (H), the lowest decile (L), and a long-short portfolio (H-L) that buys the highest and sells the lowest decile. Grey shaded areas represent NBER recession periods. The dashed vertical line marks the LLaMA3-8B model knowledge cutoff. The return time series spans from January 2005 to June 2024.

Figure 3: Report-based Forecasts and Abnormal Stock Returns: Cross-sectional Analysis



Note: This figure presents DGTW characteristic-adjusted buy-and-hold abnormal returns following research report releases, measured over a [0, +252] trading day window after each announcement. Reports are sorted based on two dimensions. First, they are categorized by the monthly median of the following characteristics at the end of the previous month: (a) market capitalization, (b) BM ratio, (c) idiosycratic volatility, (d) gross profitability, (e) analyst skill (measured by average historical EPS forecast accuracy), and (f) analyst experience (measured by years of coverage). Second, within each category, reports are sorted into deciles based on their predicted 12-month returns. The lines represent value-weighted abnormal returns of the stocks corresponding to these reports. The sample spans from 2005 to 2023.

Figure 4: Report-based Forecasts and Abnormal Stock Returns: Good News versus Bad News



Note: This figure presents DGTW characteristic-adjusted buy-and-hold abnormal returns following research report releases, measured over a [0, +252] trading day window after each announcement. Reports are sorted into deciles based on their predicted 12-month returns. Reports are classified into 'SUE(Small)' or 'SUE(Large)' groups based on whether the stocks' most recent earnings surprise falls below or above the monthly sample median. Analogous splits are applied to earnings forecast revisions and target price revisions. 'Downgrade' and 'Upgrade' represent the analysts' stock recommendation changes for these groups. The 'High' group consists of reports predicting returns in the highest decile within each month, while the 'Low' group comprises reports predicting returns in the lowest decile. The lines represent value-weighted abnormal returns of the stocks corresponding to these reports. The sample spans from 2005 to 2023.

Table 1: Summary Statistics of Analyst Reports

This table presents summary statistics for analyst reports covering S&P 1500 firms from 2000 to 2023. For each year, it reports the number of research reports, distinct brokerage firms, unique sell-side analysts, and the average report characteristics (number of pages, sentences, and tokens per report).

Year	Reports	Brokers	Analysts	Pages	Sentences	Tokens
2000	14224	49	555	5	56	1244
2001	22309	50	639	5	57	1256
2002	27555	55	791	5	57	1172
2003	28323	67	859	6	64	1257
2004	33431	70	969	6	66	1210
2005	39365	71	967	6	64	1167
2006	42347	70	951	6	64	1147
2007	43742	69	970	7	69	1217
2008	44676	72	1070	7	73	1288
2009	39425	83	1104	7	73	1281
2010	28998	86	1028	7	75	1312
2011	59367	86	1279	7	76	1263
2012	66576	85	1269	8	76	1205
2013	66960	78	1202	7	72	1143
2014	65748	73	1176	7	69	1086
2015	66699	75	1137	8	71	1138
2016	66891	71	1117	8	74	1170
2017	66846	67	1010	9	78	1234
2018	64310	62	930	9	82	1253
2019	64271	60	972	9	83	1287
2020	64567	63	964	10	85	1297
2021	56793	61	971	9	87	1300
2022	57836	60	981	10	92	1399
2023	63071	59	972	10	91	1353

Table 2: Return Predictability

This table reports the results from regressions of the specification $RET_i = \alpha_t + \beta' x_{i,t} + \varepsilon_{i,t+12}$, where $RET_{i,12m}$ is the future realized return of stock i. $x_{i,t}$ represents analyst information from each report. $\widehat{RET}_{12m}$ is the predicted next 12 months' stock return from ridge regressions. REC_{rev} denotes recommendation revision, calculated as the current report's recommendation minus the last recommendation in I/B/E/S issued by the same analyst for the same stock. EF_{rev} refers to earnings forecast revision, calculated as the current report's EPS forecast minus the last EPS forecast in I/B/E/S issued by the same analyst for the same stock, scaled by the stock price 50 days before the report release. TP_{rev} represents target price revision, calculated as the current report's target price minus the last target price in I/B/E/S issued by the same analyst for the same stock, scaled by the stock price 50 days before the report release. $Tone$ assesses the report content sentiment. I include analyst-year-month fixed effect and industry-year-month fixed effect. Standard errors are two-way clustered by firm and year-month. The t-statistics are shown in parentheses, with ***, **, and * indicating statistical significance at the 1%, 5%, and 10% level, respectively. The sample spans from 2005 to 2023.

Panel A: Forecast vs. Realised Returns Across Horizons

	RET_{1m}	RET_{3m}	RET_{6m}	RET_{9m}	RET_{12m}	RET_{18m}	RET_{24m}
$\widehat{RET}_{12m}$	0.003*	0.012***	0.026***	0.043***	0.060***	0.091***	0.114***
	(1.85)	(3.64)	(4.85)	(4.75)	(4.37)	(5.17)	(5.24)
Ana × YM FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind × YM FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.534	0.578	0.574	0.562	0.491	0.524	0.528
N	488151	488151	488151	488151	488151	473830	459623

Panel B: Numerical versus Textual Signals

	RET_{12m} (1)	RET_{12m} (2)	RET_{12m} (3)	RET_{12m} (4)	RET_{12m} (5)	RET_{12m} (6)	RET_{12m} (7)
$\widehat{RET}_{12m}$					0.087***	0.100***	0.070***
					(3.57)	(3.34)	(4.87)
REC_{rev}	-0.001					-0.005	-0.001
	(-0.18)					(-0.59)	(-0.07)
EF_{rev}		-0.351				-0.271	0.047
		(-0.90)				(-0.59)	(0.11)
TP_{rev}			0.006			0.039	0.022
			(0.25)			(1.17)	(1.09)
Tone				-0.009		-0.008	-0.013
				(-1.22)		(-0.89)	(-1.61)
YM FE	No	No	No	No	Yes	Yes	No
Ana × YM FE	Yes	Yes	Yes	Yes	No	No	Yes
Ind × YM FE	Yes	Yes	Yes	Yes	No	No	Yes
Adjusted R^2	0.489	0.573	0.474	0.490	0.162	0.165	0.579
N	480523	344742	341049	487304	837233	286129	247676

Table 3: Portfolio Statistics

This table presents the performance of value-weighted decile portfolios sorted by ridge models' return forecasts, which are based on the past {LB} months of analyst report information. The models predict 12-month ahead stock returns. For each portfolio, I report several performance measures: excess return mean and standard deviation, Sharpe ratio, α relative to Fama and French (2018) six factors, and the corresponding t-statistics for α. Portfolios are rebalanced monthly using the average of the most recent 12 months of out-of-sample predictions. The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) and a short position in the lowest decile (Low). The sample period extends from January 2005 to June 2024.

	LB = 9 months					LB = 12 months				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.53	4.68	0.39	-0.09	-0.45	0.52	4.90	0.37	-0.13	-0.68
2	0.74	4.76	0.54	0.06	0.56	0.63	4.53	0.48	-0.05	-0.44
3	0.65	4.34	0.52	-0.11	-1.38	0.66	4.53	0.50	-0.15	-1.56
4	0.71	4.54	0.54	-0.05	-0.53	0.77	4.45	0.60	0.04	0.39
5	0.75	4.71	0.55	-0.05	-0.60	0.75	4.53	0.58	-0.04	-0.55
6	0.72	4.53	0.55	0.00	0.02	0.84	4.76	0.61	0.07	0.62
7	0.81	4.89	0.57	0.04	0.30	0.73	4.75	0.53	-0.06	-0.58
8	0.95	4.68	0.70	0.13	1.61	1.07	4.83	0.76	0.25	2.07
9	0.94	5.32	0.61	0.06	0.34	0.78	5.32	0.51	-0.08	-0.61
High (H)	1.52	6.38	0.82	0.50	2.66	1.56	6.29	0.86	0.55	2.73
H-L	0.98	5.20	0.66	0.58	2.48	1.04	5.21	0.69	0.68	2.64
	LB = 18 months					LB = 24 months				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.56	4.93	0.40	-0.12	-0.54	0.53	4.98	0.37	-0.16	-0.73
2	0.56	4.56	0.43	-0.11	-0.99	0.49	4.62	0.37	-0.21	-1.98
3	0.70	4.42	0.55	-0.05	-0.74	0.71	4.44	0.55	-0.01	-0.08
4	0.69	4.53	0.53	-0.07	-0.84	0.76	4.28	0.62	0.03	0.30
5	0.77	4.43	0.60	0.04	0.36	0.58	4.75	0.42	-0.19	-1.36
6	0.84	4.69	0.62	-0.00	-0.06	0.97	4.47	0.75	0.14	1.46
7	0.69	4.63	0.52	-0.09	-0.63	0.78	4.58	0.59	-0.02	-0.13
8	1.09	4.82	0.79	0.26	2.02	0.92	5.11	0.63	0.10	0.97
9	1.12	5.56	0.70	0.22	1.42	0.90	5.17	0.60	-0.00	-0.04
High (H)	1.43	6.21	0.80	0.40	2.20	1.70	6.48	0.91	0.63	2.67
H-L	0.87	4.84	0.62	0.52	1.99	1.16	5.13	0.79	0.79	2.48

Table 4: Incremental Investment Value

This table examines the incremental investment value of analyst reports compared to existing factors. The analysis includes: full report content (RP), 18 analyst-based factors from Chen and Zimmermann (2022) (ANA), and 92 fundamental-based factors from Gu et al. (2020) (ANOM). For each strategy and combination, I report mean returns, Sharpe ratio (SR), and information ratio (IR). Newey–West (HAC) t-statistics with 12 lags for the average returns are shown in parentheses. I also report the alphas of long-short portfolios relative to four factor models: Fama and French (2015) five factors (α_{F5}), Fama and French (2018) six factors (α_{F6}), Hou et al. (2015) factors (α_{HXZ}), and Daniel et al. (2020) factors (α_{DHS}). Panel A presents individual factor performances. Panel B evaluates combinations of full report content with existing factors. The sample spans from January 2005 to December 2023.

	Mean	SR	α_{F5}	α_{F6}	α_{HXZ}	α_{DHS}	IR
Panel A: Factor Performances							
RP	1.07	0.71	0.73	0.75	0.86	1.21	-
	(3.07)	(3.04)	(2.73)	(2.86)	(3.52)	(3.90)	-
ANA	0.27	0.48	0.28	0.25	0.22	0.16	-
	(2.08)	(2.07)	(2.80)	(3.01)	(2.08)	(1.26)	-
ANOM	1.34	1.03	1.40	1.31	1.28	1.15	-
	(4.47)	(4.38)	(5.05)	(4.65)	(4.43)	(3.84)	-
ANA + ANOM	0.80	1.02	0.84	0.78	0.75	0.66	-
	(4.44)	(4.35)	(5.30)	(5.16)	(4.36)	(3.58)	-
Panel B: Report versus Factors							
RP + ANA	0.67	0.90	0.51	0.50	0.54	0.68	0.73
	(3.91)	(3.85)	(3.94)	(3.92)	(4.81)	(5.34)	(2.81)
RP + ANOM	1.21	1.55	1.06	1.03	1.07	1.18	1.16
	(6.75)	(6.44)	(7.79)	(7.37)	(7.13)	(8.01)	(3.91)
RP + ANA + ANOM	0.89	1.60	0.80	0.77	0.78	0.84	1.23
	(6.97)	(6.62)	(8.35)	(8.17)	(7.32)	(8.66)	(4.05)

Table 5: Post-Knowledge Cutoff Portfolio Performance

This table presents value-weighted decile portfolio performance during post-language model training knowledge cutoff periods, comparing four different models: BERT, RoBERTa, LLaMA2-13B, and LLaMA3-8B. Each portfolio is constructed using the corresponding model's embeddings. For each portfolio, I report three key measures: mean and standard deviation of returns, and the Sharpe ratio (SR). The analysis spans different time periods based on each model's knowledge cutoff: BERT (January 2019 to June 2024), RoBERTa (March 2019 to June 2024), LLaMA2 (October 2022 to June 2024), and LLaMA3 (April 2023 to June 2024). The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) while shorting the lowest decile (Low).

	BERT			RoBERTa			LLaMA2			LLaMA3		
	Mean	SD	SR	Mean	SD	SR	Mean	SD	SR	Mean	SD	SR
Low (L)	0.94	5.30	0.61	0.63	5.16	0.42	1.67	5.03	1.15	1.13	4.06	0.96
2	0.95	5.53	0.60	0.78	5.71	0.48	0.88	3.65	0.83	0.63	3.88	0.57
3	1.12	5.31	0.73	1.02	5.02	0.71	1.01	4.61	0.76	1.88	3.56	1.84
4	1.00	4.93	0.70	0.78	4.97	0.54	1.25	4.52	0.96	0.93	4.94	0.65
5	1.30	5.02	0.89	0.76	5.36	0.49	1.15	4.91	0.81	2.32	5.48	1.47
6	1.27	5.17	0.85	0.90	5.52	0.57	1.42	5.02	0.98	2.14	4.48	1.65
7	1.09	5.25	0.72	1.25	5.35	0.81	1.66	4.74	1.22	1.00	5.53	0.63
8	1.15	5.47	0.73	1.53	5.66	0.93	1.56	5.22	1.04	0.75	5.15	0.51
9	1.53	5.92	0.89	1.89	6.11	1.07	3.10	5.09	2.11	1.44	5.75	0.87
High (H)	2.70	7.41	1.26	2.65	7.93	1.16	3.26	7.08	1.60	3.83	7.71	1.72
H-L	1.76	6.06	1.01	2.01	6.67	1.05	1.59	7.39	0.74	2.70	7.30	1.28

Table 6: Report-based Forecasts and Characteristics

This table compares observations in the top ("High $\widehat{\text{Ret}}_{12m}$") and bottom ("Low $\widehat{\text{Ret}}_{12m}$") deciles of the monthly text-based 12-month return forecast. Panel A reports analyst attributes; Panel B reports firm characteristics. Means and medians are computed within deciles each month and then averaged over time. p-values for differences in means (medians) are from two-sample t-tests (Wilcoxon rank-sum tests). Analyst characteristics include coverage breadth, experience, broker size, and the Top10 indicator comes from I/B/E/S. Stock returns are from CRSP. Financial variables are from Compustat. See Table A1 in the Appendix for detailed variable definitions. The sample spans from January 2005 to December 2023.

Panel A: Analyst Characteristics

	High $\widehat{Ret}_{12m}$		Low $\widehat{Ret}_{12m}$			
Variable	Mean	Median	Mean	Median	Difference p-Value Means	Difference p-Value Medians
No of Industry	3.45	3.00	3.50	3.00	0.005	0.000
No of Firms	19.19	18.00	21.37	19.00	0.000	0.000
Industry Experience	16.37	16.00	16.61	16.00	0.000	0.000
Top10	0.32	0.00	0.33	0.00	0.000	0.000
Broker Size	68.61	48.00	73.41	55.00	0.000	0.000

Panel B: Firm Characteristics

	High $\widehat{Ret}_{12m}$		Low $\widehat{Ret}_{12m}$			
Variable	Mean	Median	Mean	Median	Difference p-Value Means	Difference p-Value Medians
Log Firm Size	16.18	16.15	16.06	16.05	0.000	0.000
BM	0.37	0.27	0.45	0.35	0.000	0.000
Gross Profit	0.39	0.35	0.35	0.31	0.000	0.000
Investment	1.01	0.99	0.98	0.97	0.000	0.000
Idiosyncratic Volatility	0.02	0.01	0.01	0.01	0.000	0.000

Table 7: Factor Regressions

This table reports factor regressions of value-weighted long–short (L–S) and long-only portfolios sorting on report-based return forecasts. Alphas and factor loadings are from regressions on the Fama and French (2018) six factors. t-statistics (in parentheses) are Newey–West adjusted with 12 lags. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The sample spans from January 2005 to December 2023.

	L-S	Long only
Alpha	0.739***	0.579***
	(2.77)	(2.74)
MKT-RF	0.282***	1.201***
	(4.09)	(22.77)
SMB	0.077	−0.032
	(0.53)	(−0.35)
HML	−0.720***	−0.359***
	(−4.90)	(−4.02)
RMW	0.293	0.105
	(1.46)	(1.02)
CMA	−0.829***	−0.448**
	(−3.31)	(−2.48)
MOM	−0.126	−0.046
	(−1.49)	(−0.51)
Adj. R^2	0.464	0.803

Table 8: Sources of Investment Value

This table presents an analysis of analyst report topics across five categories, examining both their distribution and investment value contribution. The categories cover Financial Analysis, Company and Industry Overview, Strategic Outlook, Risk and Governance, and Additional Content. Panel A displays aggregate statistics on topic distribution, showing the total number of sentences (in millions) and tokens (in billions) per category, along with their respective percentages of the total content. Panel B quantifies each category's contribution to portfolio performance through Shapley value decomposition, specifically examining both the Sharpe ratios and returns of the value-weighted 'H-L' portfolio.

Panel A: Topic Distribution				
Category	#Sentences	#Tokens	%Sentences	%Tokens
Financial Analysis	19.14	0.61	36.56	36.57
Company and Industry Overview	14.94	0.48	28.53	28.75
Strategic Outlook	7.92	0.26	15.13	15.61
Risk and Governance	7.40	0.24	14.14	14.35
Additional Content	2.95	0.08	5.63	4.71
Panel B: Shapley Value Decomposition				
Category	SHAP(SR)	SHAP(Ret)	%SHAP(SR)	%SHAP(Ret)
Strategic Outlook	0.24	0.28	41.34	31.43
Company and Industry Overview	0.16	0.24	27.61	26.92
Risk and Governance	0.06	0.19	11.21	21.36
Financial Analysis	0.09	0.18	16.39	19.53
Additional Content	0.02	0.01	3.44	0.77

Table 9: Sources of Investment Value in Strategic Outlook

This table analyzes analysts' strategic outlook discussions across three dimensions, quantifying their contribution to investment value through Shapley analysis. Panel A examines the temporal dimension, decomposing sentences into long-term, short-term, and combined timeframe categories. Panel B classifies sentences by sentiment orientation (negative, neutral, or positive), while Panel C categorizes content based on discussion focus (risk versus fundamental analysis). For each dimension and category, I report both absolute Shapley values (SHAP(SR) and SHAP(Ret)) and their relative percentages, measuring their contributions to portfolio Sharpe ratio and returns, respectively.

Dimension	SHAP(SR)	SHAP(Ret)	%SHAP(SR)	%SHAP(Ret)
Panel A: Timeframe				
Long-term	0.39	0.52	49.51	50.20
Short-term	0.21	0.28	26.69	26.70
Both	0.19	0.24	23.80	23.10
Panel B: Sentiment				
Negative	0.19	0.21	24.34	20.23
Neutral	0.20	0.30	25.72	28.91
Positive	0.39	0.53	49.94	50.86
Panel C: Focus				
Risk	0.10	0.13	13.00	12.25
Fundamental	0.68	0.92	87.00	87.75

Table 10: Category Portfolio Statistics

This table reports the value-weighted decile portfolio performances sorted by ridge models' forecast using the past 12 months' analyst report category-specific content. Categories include strategic outlook, company and industry overview, financial analysis, and risk and governance. The prediction target is the next 12 months' return of the stock. For each portfolio, I report several performance measures: excess return mean and standard deviation, Sharpe ratio, α relative to Fama and French (2018) six factors, and the corresponding t-statistics for α. Portfolios are rebalanced monthly using the average of the most recent 12 months of out-of-sample predictions. The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) and a short position in the lowest decile (Low). The sample period extends from January 2005 to June 2024.

	Strategic Outlook					Company and Industry Overview				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.46	4.32	0.37	-0.29	-2.18	0.76	5.69	0.46	0.08	0.38
2	0.81	4.26	0.66	0.10	1.03	0.56	4.57	0.43	-0.14	-1.15
3	0.73	4.85	0.52	-0.00	-0.03	0.61	4.28	0.49	-0.09	-1.25
4	0.71	4.62	0.54	-0.10	-1.04	0.59	4.37	0.47	-0.15	-2.08
5	0.58	4.94	0.40	-0.19	-1.43	0.75	4.54	0.57	-0.02	-0.20
6	0.68	4.84	0.49	-0.06	-0.69	0.79	4.83	0.56	0.01	0.14
7	0.79	4.92	0.55	-0.00	-0.03	0.97	4.67	0.72	0.17	1.50
8	0.87	5.07	0.59	0.08	0.63	0.81	4.74	0.59	-0.06	-0.46
9	1.08	4.92	0.76	0.25	1.74	0.94	4.79	0.68	0.08	0.85
High (H)	1.87	6.65	0.97	0.72	2.85	1.33	5.97	0.77	0.29	1.71
H-L	1.41	5.25	0.93	1.00	3.18	0.57	5.23	0.38	0.21	0.92
	Financial Analysis					Risk and Governance				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.82	5.03	0.56	0.14	0.66	0.85	4.97	0.59	0.20	1.01
2	0.53	4.26	0.43	-0.17	-1.71	0.63	4.50	0.48	-0.01	-0.10
3	0.52	4.60	0.39	-0.24	-2.38	0.49	4.49	0.38	-0.23	-2.63
4	0.73	4.65	0.54	0.00	0.03	0.77	4.69	0.57	0.02	0.23
5	0.63	4.52	0.48	-0.12	-1.49	0.70	4.65	0.52	-0.05	-0.59
6	0.93	4.62	0.70	0.19	1.59	0.82	4.55	0.63	0.04	0.41
7	0.81	4.80	0.58	-0.03	-0.28	0.72	4.77	0.52	-0.11	-1.34
8	0.75	4.82	0.54	-0.08	-0.80	0.94	4.82	0.68	0.09	0.84
9	1.07	4.99	0.74	0.17	1.33	1.06	4.89	0.75	0.15	1.34
High (H)	1.39	6.09	0.79	0.34	1.71	1.37	5.90	0.81	0.32	1.97
H-L	0.58	4.81	0.41	0.19	0.74	0.52	4.90	0.37	0.12	0.44

Table 11: Sentiment Portfolio Statistics

This table reports the value-weighted decile portfolio performances sorted by the average of four measures of the past 12 months' analyst reports. Contemporaneous Return denotes two-day DGTW characteristic-adjusted buy-and-hold abnormal returns following report release dates. Recommendation Revision denotes recommendation revision, calculated as the current report's recommendation minus the last recommendation in I/B/E/S issued by the same analyst for the same stock. Headline Sentiment represents the sentiment score of report headlines measured using a fine-tuned BERT model. Body Sentiment represents the average sentiment score of report body content measured using a fine-tuned BERT model. For each portfolio, I report several performance measures: excess return mean and standard deviation, Sharpe ratio, α relative to Fama and French (2018) six factors, and the corresponding t-statistics for α. Portfolios are rebalanced monthly. The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) and a short position in the lowest decile (Low). The sample period extends from January 2005 to June 2024.

	Contemporaneous Return					Recommendation Revision				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	1.10	6.66	0.57	0.26	1.42	0.84	4.70	0.62	0.01	0.11
2	0.93	5.99	0.54	0.14	0.86	0.87	4.81	0.62	0.09	0.95
3	0.71	5.19	0.47	-0.01	-0.10	0.88	4.61	0.66	0.10	0.75
4	0.82	4.45	0.64	0.11	1.21	0.46	4.30	0.37	-0.33	-3.63
5	0.80	4.53	0.61	-0.08	-0.93	0.54	4.90	0.38	-0.20	-1.73
6	0.79	4.41	0.62	0.00	0.03	0.69	5.31	0.45	-0.12	-1.07
7	0.84	4.56	0.64	0.00	0.05	0.81	5.61	0.50	-0.04	-0.29
8	0.94	4.78	0.68	0.08	0.75	1.02	4.90	0.72	0.15	1.32
9	1.12	5.43	0.72	0.27	2.02	0.77	4.54	0.59	0.01	0.08
High (H)	0.78	6.06	0.44	-0.10	-0.72	1.11	4.73	0.81	0.25	1.94
H-L	-0.33	4.20	-0.27	-0.36	-1.57	0.27	2.68	0.35	0.24	1.56
	Headline Sentiment					Body Sentiment				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.48	5.57	0.30	-0.23	-2.02	0.98	5.75	0.59	0.22	1.25
2	0.63	5.60	0.39	-0.09	-0.97	0.48	5.47	0.30	-0.17	-1.83
3	0.81	5.16	0.55	0.07	0.60	0.84	5.18	0.56	0.14	1.41
4	0.98	4.58	0.74	0.25	2.65	0.65	4.80	0.47	-0.09	-0.88
5	1.02	4.57	0.78	0.18	1.84	1.02	4.96	0.71	0.18	1.63
6	0.80	4.55	0.61	-0.02	-0.17	0.87	4.71	0.64	0.05	0.41
7	0.78	4.77	0.57	-0.08	-0.97	0.81	4.75	0.59	-0.10	-1.35
8	0.67	4.55	0.51	-0.19	-1.68	0.82	4.59	0.62	-0.03	-0.31
9	0.85	4.93	0.60	-0.03	-0.19	0.90	4.50	0.70	0.03	0.29
High (H)	0.74	5.11	0.50	-0.15	-1.50	0.82	4.56	0.63	-0.01	-0.16
H-L	0.27	3.31	0.28	0.08	0.58	-0.16	4.17	-0.13	-0.24	-0.97

A Additional Tables and Figures

(1) Company and Industry Overview

(2) Financial Analysis

(3) Strategic Outlook

(4) Risk and Governance

Figure A1: Word Clouds of Topics

This figure presents word clouds for 4 categories frequently discussed in analyst reports. Each word cloud visually represents the most common terms associated with the topic, with word size indicating term frequency. The categories include Company and Industry Overview, Financial Analysis, Strategic Outlook and Risk and Governance. The "Additional Content" category is excluded from the visualization.

Table A1: Variables Description

This table shows the definitions of numerical measures.

Numerical Measures	Definition and/or sources
Panel A: Firm-Level	
Log Firm Size	The logarithm of market value equity of the firm ($CSHOQ * PRCCQ$) at the end of the month prior to report release.
BM	The book value of equity ($SEQ + TXDB + ITCB - PREF$) scaled by the market value of equity ($CSHOQ * PRCCQ$) t the end of the month prior to the report release.
SUE	Earnings surprise, calculated as the actual EPS minus the last consensus EPS forecast before the earnings announcement. Consensus EPS is the median value of 1-year EPS forecast within a 90-day window of all analysts following the firm. The unexpected earnings is scaled by price per share at the fiscal quarter end.
Gross Profit	Revenue (sale) - cost of goods solds (cogs), divided by total assets (at).
Investment	Ratio of capital investment (capx) to revenue (revt) divided by the firm-specific 36-month rolling mean of that ratio. Exclude if revenue less than \$10m.
Idiosyncratic Volatility	Standard deviation of residuals from Fama-French three-factor regressions using the past month of daily data.
Panel B: Analyst-Level	
No of Industry	The number of Fama–French 48 industries covered by the analyst at year t.
No of Firms	The number of firms followed by the analyst at year t.
Industry Experience	Analyst-year measure equal to the number of calendar years since the analyst's first appearance in I/B/E/S.
Top10	Indicator equal to 1 if a brokerage ranks in the top ten by Broker Size within year t, and 0 otherwise. (Ranking is computed cross-sectionally each year.)
Broker Size	The number of distinct analysts employed by a brokerage in a calendar year, proxied by the count of unique analyst–broker pairs with at least one EPS forecast in I/B/E/S that year
Panel C: Report-Level	
REC_{REV}	Recommendation revision, calculated as the current report's recommendation minus the last recommendation in I/B/E/S issued by the same analyst for the same firm.
EF_{REV}	Earnings forecast revision, calculated as the current report's EPS forecast minus the last EPS forecast in I/B/E/S issued by the same analyst for the same firm, scaled by the stock price 50 days before the report release.
TP_{REV}	Target price revision, calculated as the current report's target price minus the last target price in I/B/E/S issued by the same analyst for the same firm, scaled by the stock price 50 days before the report release.
Tone	Net sentiment of the report text, computed by classifying each sentence with a fine-tuned BERT model as positive, negative, or neutral; defined at the report level as Tone $_{jt} = \left(N_{jt}^{+} - N_{jt}^{-}\right)/N_{jt}$, where $N_{jt}^{+}$and $N_{jt}^{-}$are counts of positive and negative sentences and N_{jt} is the total number of sentences.

Table A2: Fundamental-based Predictors

This table reports the market and firm-related fundamental factors considered in this paper. I follow Gu et al. (2020) to construct the dataset. Gu et al. (2020) provide a detailed description of the dataset in their appendix.

Acronym	Description	Acronym	Description
absacc	Absolute accruals	mom36m	36-month momentum
acc	Working capital accruals	mom6m	6-month momentum
aeavol	Abnormal earnings announcement volume	ms	Financial statement score
age	# years since first Compustat coverage	mvel1	Size
agr	Asset growth	mve_ia	Industry-adjusted size
baspread	Bid-ask spread	niincr	Number of earnings increases
beta	Beta	operprof	Operating profitability
betasq	Beta squared	orgcap	Organizational capital
bm	Book-to-market	pchcapx_ia	Industry adjusted % change in capital expenditures
bm_ia	Industry-adjusted book to market	pchcurrat	% change in current ratio
cash	Cash holdings	pchdepr	% change in depreciation
cashdebt	Cash flow to debt	pchgm_pchsale	% change in gross margin - % change in sales
cashpr	Cash productivity	pchquick	% change in quick ratio
cfp	Cash flow to price ratio	pchsale_pchinvt	% change in sales - % change in inventory
cfp_ia	Industry-adjusted cash flow to price ratio	pchsale_pchrect	% change in sales - % change in A/R
chatoia	Industry-adjusted change in asset turnover	pchsale_pchxsga	% change in sales - % change in SG&A
chcsho	Change in shares outstanding	pchsaleinv	% change in sales-to-inventory
chempia	Industry-adjusted change in employees	pctacc	Percent accruals
chinv	Change in inventory	pricedelay	Price delay
chmom	Change in 6-month momentum	ps	Financial statements score
chpmia	Industry-adjusted change in profit margin	quick	Quick ratio
chtx	Change in tax expense	rd	R&D increase
cinvest	Corporate investment	rd_mve	R&D to market capitalization
convind	Convertible debt indicator	rd_sale	R&D to sales
currat	Current ratio	realestate	Real estate holdings
depr	Depreciation / PP&E	retvol	Return volatility
divi	Dividend initiation	roaq	Return on assets
divo	Dividend omission	roavol	Earnings volatility
dolvol	Dollar trading volume	roeq	Return on equity
dy	Dividend to price	roic	Return on invested capital
ear	Earnings announcement return	rsup	Revenue surprise
egr	Growth in common shareholder equity	salecash	Sales to cash
ep	Earnings to price	saleinv	Sales to inventory
gma	Gross profitability	salerec	Sales to receivables
grCAPX	Growth in capital expenditures	secured	Secured debt
grltnoa	Growth in long term net operating assets	securedind	Secured debt indicator
herf	Industry sales concentration	sgr	Sales growth
hire	Employee growth rate	sin	Sin stocks
idiovol	Idiosyncratic return volatility	sp	Sales to price
ill	Illiquidity	std_dolvol	Volatility of liquidity (dollar trading volume)
indmom	Industry momentum	std_turn	Volatility of liquidity (share turnover)
invest	Capital expenditures and inventory	stdacc	Accrual volatility
lev	Leverage	stdcf	Cash flow volatility
lgr	Growth in long-term debt	tang	Debt capacity/firm tangibility
maxret	Maximum daily return	tb	Tax income to book income
mom12m	12-month momentum	turn	Share turnover
mom1m	1-month momentum	zerotrade	Zero trading days

Table A3: Analyst-based Predictors

This table reports the analyst factors considered in this paper. I compile the data from Chen and Zimmermann (2022) website.

Acronym	Journal (Publish Year)	Description
AnalystRevision	FAJ (1984)	EPS forecast revision
AnalystValue	JAE (1998)	Analyst Value
AOP	JAE (1998)	Analyst Optimism
ChangeInRecommendation	JF (2004)	Change in recommendation
ChForecastAccrual	RAS (2004)	Change in Forecast and Accrual
ChNAnalyst	ROF (2008)	Decline in Analyst Coverage
ConsRecomm	JF (2001)	Consensus Recommendation
CredRatDG	JF (2001)	Credit Rating Downgrade
DownRecomm	JF (2001)	Down forecast EPS
EarningsForecastDisparity	JFE (2011)	Long-vs-short EPS forecasts
ExclExp	RAS (2003)	Excluded Expenses
FEPS	WP (2006)	Analyst earnings per share
fgr5yrLag	JF (1996)	Long-term EPS forecast
ForecastDispersion	JF (2002)	EPS Forecast Dispersion
Recomm_ShortInterest	AR (2011)	Analyst Recommendations and Short-Interest
REV6	JF (1996)	Earnings forecast revisions
sfe	AR (2001)	Earnings Forecast to price
UpRecomm	JF (2001)	Up Forecast

Table A4: Topic Categories and Descriptions

Topic	Desriptions
Company and Industry Overview	
Executive Summary	Provides a high-level overview of the report's key findings and conclusions; includes a brief description of the company, its industry, and the purpose of the report; highlights the most important points from the analysis, such as the company's financial performance, competitive position, and growth prospects.
Company Overview	Offers a comprehensive description of the company, including its history, business model, and key products or services; discusses the company's organizational structure, management team, and corporate governance; analyzes the company's mission, vision, and strategic objectives.
Industry Analysis	Provides an in-depth analysis of the industry in which the company operates; includes information on market size, growth trends, and key drivers; discusses the regulatory environment, technological advancements, and other external factors affecting the industry; analyzes the industry's competitive dynamics and the company's position within the industry.
Competitive Landscape	Identifies the company's main competitors and their market share; compares the company's products, services, and pricing strategies with those of its competitors; analyzes the strengths and weaknesses of the company and its competitors; discusses potential new entrants and substitutes that could disrupt the competitive landscape.
Business Segments	Provides a detailed analysis of the company's various business segments or divisions; discusses the financial performance, growth prospects, and challenges of each segment; analyzes the contribution of each segment to the company's overall revenue and profitability.
Growth Strategies	Discusses the company's strategies for driving future growth, such as organic growth initiatives, product innovations, and geographic expansions; analyzes the company's mergers and acquisitions (M&A) strategy and potential targets; examines the company's investments in research and development (R&D) and marketing.
Financial Analysis	
Income Statement Analysis	Analyzes the company's revenue, expenses, and profitability.
Balance Sheet Analysis	Examines the company's assets, liabilities, and shareholders' equity.
Cash Flow Analysis	Analyzes the company's cash inflows and outflows to evaluate liquidity.
Financial Ratios	Discusses key ratios like profitability, liquidity, and solvency ratios.
Strategic Outlook	
Investment Thesis	Summarizes the key reasons for investing (or not investing) in the company's shares; discusses the potential catalysts and risks that could impact the company's valuation and stock price performance; provides a target price or price range for the company's shares based on the valuation analyses and investment thesis.
Valuation	Estimates the intrinsic value of the company's shares using various valuation methodologies, such as discounted cash flow (DCF) analysis, relative valuation multiples, and sum-of-the-parts analysis; compares the company's valuation with that of its peers and historical benchmarks; discusses the key assumptions and sensitivities underlying the valuation analyses.
Risk and Governance	
Risk Factors	Identifies and analyzes the key risks facing the company, such as market risks, operational risks, financial risks, and legal/regulatory risks; discusses the potential impact of these risks on the company's financial performance and growth prospects; examines the company's risk management strategies and mitigation measures.

Table A4 – continued from previous page

Topic	**Desriptions**
Management and Governance	Provides an overview of the company's management team, including their experience, expertise, and track record; analyzes the company's corporate governance practices, such as board composition, executive compensation, and shareholder rights; discusses the company's succession planning and key person risks.
ESG	Analyzes the company's performance and initiatives related to environmental sustainability, social responsibility, and corporate governance; discusses the potential impact of ESG factors on the company's reputation, risk profile, and financial performance; examines the company's compliance with ESG regulations and industry standards.
Additional Content	
Appendices and Disclosures	Includes additional supporting materials, such as financial statements, ratio calculations, and detailed segment data; provides important disclosures, such as the analyst's rating system, potential conflicts of interest, and disclaimers; discusses the limitations and uncertainties of the analysis and the need for further due diligence by investors.
None of the Above	Covers any content that does not fall into the specified topics.

Table A5: Incremental Investment Value: Strategic Outlook

This table examines the incremental investment value of analyst reports compared to existing factors. The analysis includes: strategic outlook content (SO), 18 analyst-based factors from Chen and Zimmermann (2022) (ANA), and 92 fundamental-based factors from Gu et al. (2020) (ANOM). For each strategy and combination, I report mean returns, Sharpe ratio (SR), and information ratio (IR), with corresponding t-statistics in parentheses. I also report the alphas of long-short portfolios relative to four factor models: Fama and French (2015) five factors (α_{F5}), Fama and French (2018) six factors (α_{F6}), Hou et al. (2015) factors (α_{HXZ}), and Daniel et al. (2020) factors (α_{DHS}). Panel A presents individual factor performances. Panel B evaluates combinations of strategic outlook content with existing factors. The sample spans from January 2005 to December 2023.

	Mean	SR	α_{F5}	α_{F6}	α_{HXZ}	α_{DHS}	IR
Panel A: Factor Performances							
SO	1.30	0.87	0.95	0.96	1.04	1.46	-
	(3.79)	(3.74)	(3.03)	(3.01)	(3.50)	(3.89)	-
ANA	0.27	0.48	0.28	0.25	0.22	0.16	-
	(2.08)	(2.07)	(2.80)	(3.01)	(2.08)	(1.26)	-
ANOM	1.34	1.03	1.40	1.31	1.28	1.15	-
	(4.47)	(4.38)	(5.05)	(4.65)	(4.43)	(3.84)	-
ANA + ANOM	0.80	1.02	0.84	0.78	0.75	0.66	-
	(4.44)	(4.35)	(5.30)	(5.16)	(4.36)	(3.58)	-
Panel B: Strategic Outlook versus Factors							
SO + ANA	0.78	1.07	0.62	0.61	0.63	0.81	0.96
	(4.68)	(4.57)	(4.34)	(4.26)	(5.30)	(4.95)	(3.45)
SO + ANOM	1.32	1.69	1.17	1.14	1.16	1.30	1.34
	(7.37)	(6.97)	(7.92)	(7.79)	(8.21)	(7.92)	(4.24)
SO + ANA + ANOM	0.97	1.74	0.88	0.84	0.84	0.92	1.41
	(7.57)	(7.14)	(9.05)	(9.55)	(9.51)	(8.64)	(4.35)

Table A6: Post-Knowledge Cutoff Portfolio Performance: Strategic Outlook

This table presents value-weighted decile portfolio performance during post-language model training knowledge cutoff periods, comparing four different models: BERT, RoBERTa, LLaMA2-13B, and LLaMA3-8B. Each portfolio is constructed using the corresponding model's embeddings of strategic outlook content. For each portfolio, I report three key measures: mean and standard deviation of returns, and the Sharpe ratio (SR). The analysis spans different time periods based on each model's knowledge cutoff: BERT (January 2019 to June 2024), RoBERTa (March 2019 to June 2024), LLaMA2 (October 2022 to June 2024), and LLaMA3 (April 2023 to June 2024). The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) while shorting the lowest decile (Low).

	BERT			RoBERTa			LLaMA2			LLaMA3		
	Mean	SD	SR	Mean	SD	SR	Mean	SD	SR	Mean	SD	SR
Low (L)	0.82	5.69	0.50	0.97	5.36	0.63	1.08	4.23	0.89	0.01	3.48	0.01
2	1.10	4.92	0.78	0.77	5.04	0.53	1.21	4.37	0.96	1.19	3.77	1.09
3	0.92	5.16	0.61	0.98	5.22	0.65	1.15	4.51	0.88	1.05	4.19	0.87
4	1.39	5.04	0.95	0.99	4.96	0.69	1.39	4.49	1.07	2.05	4.97	1.43
5	1.20	5.15	0.81	0.68	5.37	0.44	1.82	5.01	1.26	1.69	3.60	1.63
6	0.66	5.52	0.42	0.95	5.34	0.62	1.91	4.85	1.36	0.66	4.56	0.51
7	1.14	5.48	0.72	1.35	5.04	0.93	1.45	4.73	1.06	0.67	5.09	0.46
8	1.53	5.77	0.92	1.34	5.56	0.84	2.35	5.64	1.44	1.77	4.81	1.28
9	2.17	6.67	1.13	1.94	7.33	0.92	2.30	6.14	1.30	2.26	5.13	1.53
High (H)	3.19	8.09	1.37	2.78	8.28	1.16	3.73	7.10	1.82	5.29	7.88	2.32
H-L	2.37	6.51	1.26	1.81	7.37	0.85	2.64	6.95	1.32	5.28	7.73	2.37